ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Bachelor Thesis

"The influence of dietary awareness on the relationship between education and health behaviour."

Author: Maurice Heijboer

Student number: 335008

Supervisor: Dr. Hans van Kippersluis

18 July 2013

Rotterdam

Table of contents

1. Summary

A lot of research has been done to study and explain the education-health gradient, which is the relationship between education and health. The higher educated appear to have a higher life expectancy and live more years in good health. Several explanations for this education-health gradient can be given from the literature: Poor health leads to low levels of education, third factors increase both health and education and higher education leads to an improved health. This thesis focuses on estimating the effect of education and specific knowledge on health measures and thereby tries to extend the existing literature.

Cutler and Lleras-Muney (2010) try to explain the relationship between education and health behaviour. They show that specific knowledge about smoking and drinking leads to a decrease of the education-health gradient in the corresponding health behaviours. Besides smoking and drinking, overweight is a major health problem in many countries. This paper investigates the question to what extent diet related knowledge influences the education-health gradient in diet related health behaviour.

To give an answer to this research question we first use a linear regression model to estimate the effect of education on diet related health behaviour. Education has a significant negative influence on BMI and a significant positive influence on the change that people find it important to choose a healthy diet and maintain a healthy weight. Consequently, questions about dietary knowledge are added to the regression in order to estimate the possible change in the effect of education on the different health behaviours. Entering the knowledge questions leads to a significant and mostly large reduction of the influence of education on these health behaviours.

Second, we use a linear regression model to estimate the effects of these health behaviours (BMI and the importance of people to choose a healthy diet and maintain a healthy weight) on an individual's state of health. BMI has a significant negative effect on the state of health, whereas the reported importance of people to choose a healthy diet and maintain a healthy weight has a significant positive effect on their state of health.

Concluding, we could state that dietary awareness has a large influence on diet related health behaviours and that these diet related health behaviours have a significant impact on the present state of health. So the education-health gradient in diet related health behaviour is to a quite large extent influenced by dietary awareness.

2. Introduction

The relation between education and health, called the education-health gradient, is well known and highly studied. In the Netherlands the higher educated have a life expectancy which is 6 to 7 years higher than the lower educated. Furthermore, they live 16 to 19 years longer in good health (CBS, 2008).There are however several explanations for this relationship between education and health, and causality is not straightforward. From the existing literature, three reasons for the existence of the education-health gradient can be given (Cutler and Lleras-Muney, 2006):

1. Poor health leads to low levels of education;
2. Third factors increase both health and education;
3. Higher education leads to an improved health.

Corresponding to these reasons there are different directions of research, each with their own hypotheses and assumptions regarding the causal pathways.

First, there is a branch of research that studies the effect of health on education. Early health conditions of children are often used to estimate later education levels. Behrman and Rosenzweig (2004) find that children with a lower birth weight obtain less education than those born with higher weights. This even applies to twins. Also, children that at later ages being more sick or underfed, are more likely to miss school, learn less when in school and eventually obtain less years of education (Case, Fertig and Paxson, 2005). There are however doubts about the explanatory potential of health on education. Cutler and Lleras-Muney (2006) state: “If this was important, one would expect the education-health gradient to be diminishing over time: very few children in the United States today are unable to attend school because of their health. But the education-health gradient is rising.”

Second, there can be unobserved third factors that increase both health and education. For example family background or other individual characteristics. Fuchs (1982) tests the relationship between time preference and schooling, health behaviours and health status. A different discount rate and thus a different value of the future could possibly explain differences in schooling and health. He however finds a weak correlation of the discount rate with schooling, health behaviours and health status. Cutler and Lleras-Muney (2006) add family background measures, time preference and other individual characteristics to their results and conclude that these measures lower the effect of education on health measures and health behaviour, but that the initial effect remains large and significant. There are no third factors known that can wipe away the relationship between education and health by explaining them both.

Third, there is a lot of research estimating the effect of education on health measures. Several natural experiments have been done to investigate whether this effect of education on health is causal. Lillard and Eamon (2010) use the introduction of compulsory schooling laws in the United States to conclude that an extra year of compulsory schooling leads to a significant reduction in the probability of death within the next 10 years. Changes in education policy are used as exogenous shock to separate to effect of education on health measures. Van Kippersluis et al. (2011) exploit the introduction of a Dutch schooling law to conclude that education significantly reduces mortality in old age. Also for other countries such as Denmark (Arendt, 2005) and the United States (Lleras-Muney, 2005) effects of education on health measures are identified. Findings of other studies regarding other countries and using slightly different methodologies vary – Clark and Royer (2010) and Albouy and Lequien (2009) – and sometimes don't show significant effects. The overall result of these natural experiments is however that the relation between education and health is at least partly causal, whereas the specific explanations remain mostly uncertain.

Cutler and Lleras-Muney (2010) use different datasets to explain the relationship between education and health. They show that differences in social-economic factors and family background account for about 30% of the education-health gradient. Knowledge and other measures of cognitive ability account for another 30%. They show furthermore that each additional year of schooling is associated with a 3.0 percentage point lower probability of smoking, a 1.8 percentage point lower probability of being a heavy drinker and a 1.4 percentage point reduction of the probability of being obese. They further decompose the effect of knowledge on the education-health gradient by looking at smoking and drinking behaviour. Specific knowledge about smoking leads to a decrease of the education-health gradient in smoking, whereas knowledge about drinking has only little impact on drinking. Specific knowledge is therefore a source of differences in health behaviour.

After smoking, overweight is the leading behavioural cause of death in many countries. A conservative estimation of overweight mortality in the United States in the year 2000 is 385.000, which corresponds to 16% of total deaths (Mokdad, 2004). Cutler and Lleras-Muney (2010) show that each additional year of schooling significantly reduces the Body Mass Index and the probability of being obese. They also show that specific knowledge can lead to differences in smoking and drinking behaviour. But measuring the effect of specific knowledge, they only look at these two variables and don't look at diet related knowledge and how this possibly influences diet related behaviour. This paper extends the research regarding the effect of knowledge on the education-health gradient by looking at dietary awareness. It investigates the effect of diet related knowledge on the education-health gradient in diet related health behaviour. If we are able to conclude that

specific knowledge influences the education-health gradient in diet related health behaviour, this can lead to meaningful insights and policy implications. One of the driving forces behind the education-health gradient can be decomposed and it may also be advisable to invest in dietary knowledge of students and pupils in order to improve their future health.

The research question is: *To what extent is the education-health gradient in diet related health behaviour influenced by dietary awareness?*

Structure of the thesis

This thesis first provides a theoretical framework in order to explain different health behaviours by education. Grossman (1972) constructs a model for the demand of the commodity 'good health' and introduces the concept of 'allocative efficiency', where the higher educated produce more health because they choose a better mix of inputs, we are now able to be more concrete. Van Kippersluis and Galama (2013) provide an important extension to this model and make a distinction between healthy and unhealthy consumption. In this thesis, we test to what extent specific knowledge of this (un)healthy consumption differs across educational groups.

PSID data of the 1999th and 2001th is used in order to answer the research question. This data will be described in the data section. To analyze this data linear regression is used to estimate the effect of dietary awareness on the education-health gradient in diet related health behaviour and consequently the influence of diet related health behaviour on the present state of health. This can be found in the methodology section.

After providing the results of these analyses and then the interpretation of the results, we are able to draw conclusions and answer the research question. At last, limitations of this thesis and further recommendations will be given.

3. Theoretical Framework

From empirical research it is clear that the higher educated are healthier and have a higher life expectancy. Grossman (1972) therefore constructs a model for the demand of the commodity 'good health'. Van Kippersluis and Galama (2013) provide two important extensions of this model.

From human capital theory it is known that individuals invest in themselves in order to raise their productivity in the market sector, where monetary earnings are produced, and in the nonmarket or household sector, where commodities are produced that enter his utility function. Based on this theory Grossman (1972) constructs a model of the demand for health capital. This model is based on human capital theory, where people invest in themselves to raise their productivity in the market as well in the non-market sector. Grossman argues however that health capital differs from human capital, because the stock of health determines the total amount of time individuals can spend producing monetary earnings and commodities, not the earnings and commodities itself. Health is demanded by individuals for two reasons: as a consumption commodity and as an investment commodity. As a consumption commodity it directly enters the utility functions, people feel better when they are healthier. As an investment commodity, it determines the total amount of time available for market and nonmarket activities. An investment in an individual's stock of health delivers him more 'healthy time', allowing them to spend more time in market and nonmarket activities. The gain of this investment is therefore the monetary value of the reduction in lost (or 'unhealthy') time. It is assumed that individuals inherit an initial stock of health that depreciates over time and can be increased by investment.

In Grossmans model utility is dependent on H_t, the stock of health at age t and Z_t the total consumption of another commodity at age t. Net investment in the stock of health equals gross investment minus depreciation:

$$H_{t+1} - H_t = I_t - \delta_t H_t$$

I_t is gross investment and δ_t is the rate of depreciation of the health stock. Individuals are born with an initial health stock and by investments they can add to this stock. Gross investments in health capital have two inputs: 1) own time of the individuals and 2) market goods, such as medical care, diet and exercise. These two inputs are produced by household production functions, which depend on certain 'environmental variables' that influence the efficiency of the production process. The most important of these variables is the education level. The depreciation rate depends on age, diseases, accidents, etc.

In the model of Grossman consumers produce gross investments in health and other commodities. They do this according to a set of household production functions:

$$I_t = I_t\,(M_t, TH_t; E_i),$$

$$Z_t = Z_t(X_t, T_t;\ E_i)$$

Where M_t is medical care, TH_t is time spent in improving health, Z_t is production of a commodity with goods input X_t, T_t is the time spent on this commodity and E_i is human capital, which is measured by education in the model of Grossman. According to his model, people invest in their stock of human capital by formal schooling and on-the-job training to realize potential gains in productivity.

To determine the optimal choices of individuals it is assumed that they want to maximize their lifetime utility, they have perfect knowledge and they are able to allocate their time between the different activities. The amount of time they have is divided by labour time, leisure time and there is time lost because illness. During their leisure time they make a choice between health producing time and non-health producing time. As stated in the equations above, they choose between investments in health versus production of other commodities.

A healthy time production function can now be derived. It has an upper limit of 8,760 hours in a year and the marginal product of this capital diminishes:

$$h_t = 8760 - BH_t^{-C}$$

Now, B and C are positive constants. From this equation the marginal product of health capital can be derived:

$$G_t = \left(\frac{\delta h_t}{\delta H_t}\right) = BCH_t^{-C-1}$$

In this investment model, the consumption of health is not taken into account. The marginal utility of healthy time and the negative marginal utility of unhealthy time are therefore equal to zero. Also in this model, the health stock at equilibrium level, at constant marginal costs of gross investment in health, is determined by equalizing the marginal monetary gain of health capital to the opportunity cost of that capital equal. The return on investment in health is weighed against the investment in other commodities.

People now make an optimal choice between investment in health capital and other commodities where marginal costs (MC) equal marginal benefits (MB).

$$MC = r - \tilde{\pi}_t + \delta_t$$

Where r is the interest rate on other investments, $\tilde{\pi}_t$ is the continuously compounded percentage rate of change in marginal cost with age and δ_t is again the rate of depreciation. The latter is the opportunity cost of investments in health.

$$MB = (W_t * G_t)/\pi_t$$

Where W_t is the wage rate and π_t is the marginal cost of investments in health. Gross investments in health are assumed to be positive in this model.

Now from these equations a demand curve for health capital can be formed by:

$$\text{Ln}\, H_t = \ln BC + \epsilon \ln W_t - \epsilon \ln \pi_t - \epsilon \ln \delta_t$$

Where $\epsilon = 1/(1 + C)$ is the elasticity of the demand curve. Grossman then obtains a reduced-form demand curve for health capital, by making assumptions about the nature of the depreciation rate function and the marginal cost of the gross investment function:

$$\ln \delta_t = \delta_1 + \tilde{\delta} t$$

Where δ_1 is the rate of depreciation during some initial period and $\tilde{\delta} t$ is the continuously compounded rate of increase in the rate of depreciation with age.

Grossman then transforms the production function of gross investment to a Cobb-Douglas function and creates a marginal cost function:

$$\ln I_t = \alpha \ln M_t + (1 - \alpha) \ln T_t + \rho E$$

M_t is a market good or a vector of market goods to produce gross investments in health. T_t is the consumer's time input and E is an index of human capital. α is a parameter of the output elasticity of M_t ,$(1 - \alpha)$ is the output elasticity of T_t and ρ the percentage improvement in nonmarket productivity due to human capital.

Now, a reduced-form demand curve for health can be given by:

$$\ln H_t = \alpha\epsilon \ln W_t - \alpha\epsilon \ln P_t + \rho\epsilon E - \tilde{\delta}\epsilon t - \epsilon \ln \delta_t$$

Where P_t is the price of M_t. The equation shows that the demand for health capital is positively related to the wage rate and the stock of health capital and on the other hand negatively correlated to the price of M_t age and the depreciation rate in the initial period.

From Grossman's work it appears that when age, the wage rate, and several others variables are held constant, education has a positive effect on health. Grossman (2000) explains this by using the concept of 'productive efficiency'. An increase in education leads to an increase in the efficiency of producing health. Even with the same inputs the higher educated produce more health, because they use the inputs more efficient.

On the other hand, 'allocative efficiency' models (e.g. Kenkel, 1991) assume that the higher educated have more knowledge about the relationship between various certain inputs and health and choose a better mix of these inputs which yield more health. Education now only has an impact on health because a change in inputs. This in contrast to the productive efficiency models, where education leads to a more efficient use of inputs by the higher educated.

Where Grossman (1972) only looks at consumption as a whole and therefore uses one variable for all kinds of consumption, van Kippersluis and Galama (2013) divide consumption into two parts: healthy and unhealthy consumption. From empirical research it is known that richer people engage more in moderately unhealthy behaviour, but less in severely unhealthy behaviour. Cutler and Lleras Muney (2010) show that more affluent people smoke less and tend to drink more 'moderately', but less 'excessive' than the less wealthy. This can't be fully explained by theoretical economic models of consumption and saving, because wealth enables people to consume more. Here we see that wealthier people consume less. Van Kippersluis and Galama (2013) therefore assume that consumers take health consequences of consumption into account.

Individuals now maximize their life-time utility function

$$\int_0^T U[C_h(t), C_u(t), H(t)]e^{-\beta t}dt \,,$$

Where T denotes the life span, β is a subjective discount factor. Healthy consumption, $C_h(t)$, provides utility and lowers the biological aging rate. Unhealthy consumption, $C_u(t)$, provides utility but increases the biological aging rate. Time t is measured from the time an individual has completed her education and joined the labour force. Utility now increases with healthy consumption, unhealthy consumption and with health, but at a decreasing rate.

From the concept of 'allocative efficiency' according to the theory of Grossman, where the higher educated produce more health because they choose a better mix of inputs, we are now able to be more concrete. The depreciation rate of health capital, or biological aging rate, decreases with healthy consumption, but increases with unhealthy consumption. In other words, the health cost increases in the 'unhealthiness' of a good. If a distinction between healthy and unhealthy

consumption can be made, we can test to what extent specific knowledge of this (un)healthy consumption differs across educational groups.

4. Data

The dataset of the Panel Study of Income Dynamics (PSID) will be used in order to research the influence of dietary awareness on the education-health gradient. Directed by the Institute for Social Research at the University of Michigan, this dataset contains information from a representative sample of the United States population. It is the world's longest running household panel survey (PSID, General Guide, 2013). Interviews are held biannually to conduct information about employment, income, wealth, expenditures, health, education, marriage, childbearing, philanthropy, and numerous other topics. The majority of the interviews are conducted via telephone by using a computer-assisted telephone collection technology. Information about each family member is collected, but the questions are asked to the head of the family. If the family consists of a husband-wife pair, the husband is designated the head. The PSID collects data of families since 1968 and follows the original families, and eventually their kids, in time. As members of sample families grow up, move out, and form their own economically independent households, they are interviewed separately, increasing the overall number of interviews conducted each wave. During the past decades several validation studies have been performed in order to maintain representative. There is a close alignment of weighted estimates from PSID with those from other U.S. benchmark studies – the March Current Population Survey for income, the Survey of Consumer Finances for wealth, the National Health Interview Survey for health status and health behaviours, the Consumer Expenditures Survey for expenditures, and the American Time Use Survey for time use behaviours (McGonagle, Schoeni, Sastry, & Freedman, 2012). This supports the claim that PSID remains representative of the U.S. population. The response rate of the 1999th wave, which will be used in the analysis, is 93.1% (McGonagle, Schoeni, Sastry, & Freedman, 2012).

Since 1999 health is a core topic of the PSID. People are asked to their present state of health, other possible health problems, their health condition, health behaviours and other related factors. In the wave of 1999 also questions were asked regarding dietary awareness. Originally these questions were asked to assess awareness of governmental outreach in the area of the "food pyramid", the "5-a-day programme", and "dietary guidelines for Americans". These are all governmental guidelines and campaigns aimed to help and motivate people to become healthier by encouraging them to increase their consumption of fruits and vegetables (PSID, Data on Food and Nutrition, 2013). In the

1999th wave, also questions were asked regarding the importance of a healthy diet; these are questions about the importance of fruits and vegetables, a diet low in saturated fat, fibre in diet, grain products and a healthy weight. Questions regarding people's awareness of diet-related health problems and recommended dietary guidelines were asked as well. These are questions about the awareness (knowledge) of the relationship between a lack of certain nutrients and health problems and other diet-related programmes.

To research the influence of dietary awareness on the education-health gradient multiple variables regarding health behaviours will be used. Almost all of these variables in the PSID are self-reported. Data containing measured behaviour is however hardly available. It is possible that biases in self-reported variables vary across different health behaviours. To mitigate this possible bias several measures of health behaviour will be used in this analysis. Nevertheless, not much is known about whether biases in reporting vary systematically by, for example, education (Cutler & Lleras-Muney, 2010).

In case of no response to a question or when people answer "don't know" this observation is filtered out of the regression. Total number of observations is 4,600.

To measure education level the numbers of grades of education people have completed will be used. The values of this variable represent the actual grade of school a person has completed. It runs from 1 to 17, where 1 is the first year at primary school, 9 is the first year at high school, 13 is the first year at university and 17 is at least some postgraduate work completed. The median education grade completed is 12, which represents a high school diploma. The frequencies of the number of grades completed are roughly comparable to the nationwide educational attainment (Census Bureau, 1999). Only the highest category with at least some postgraduate work is somewhat overrepresented.

As additional explanatory variables of health behaviour will be used: the Body Mass Index (BMI), people's importance of choosing a diet with plenty of fruits and vegetables, the importance of choosing a diet low in saturated fat, the importance of choosing a diet with adequate fibre, the importance of a diet with plenty of grain products and the importance to maintain a healthy weight. The BMI variable is constructed by making use of the variables height and weight. The mean BMI of the sample is 27.4. All variables regarding the importance of diet-related factors are measured on a four-point scale, where 1 is not at all important, 4 is very important. Because we also use these variables as dependent variable later on, we change these variables to a two-point scale for reasons of comparability. We define 1, 2 and 3 as not (very) important and 4 as important. The majority of people reports it to be important to choose a diet with plenty of fruits and vegetables (54%) and to

maintain a healthy weight (63%). 49% of the people find it important to choose a diet low in saturated fat. To choose adequate fibre and plenty grain products seems to be less important, with 42% and 35% respectively.

As control variables we use age, family income, ethnicity, state of health until age 16, marital status, education level of father and education level of mother. Ideally, we control for variables that determine education but cannot be affected by it. Family income doesn't fit to this criterion, it is clearly related to the education level, therefore more endogenous and can bias the effect of education. We follow however other studies (e.g. Cutler and Lleras-Muney, 2010) and include it as a sensitivity test, because the income-health gradient can't be ignored.

Because we will look at the effect of education on health, we restrict the sample to individuals ages 25 and above, so education has been (mostly) completed. The sample has a mean age of 46. Family income consists of total taxable income, including income from assets and net profits, transfer income and social security income. This variable contains negative values because of net losses. Because this variable is positively skewed, we will change it to a positive log-scale. PSID also reports ethnicity of the respondents. This is done in six categories. 1 is white, 2 is black, 3, 4, 5, 6 and 7 are smaller, other categories. We define values 3 to 7 as one 'other' category. 68% of the respondents report themselves as being white, 25% as black and 7% as other. To control for childhood health conditions state of health until age 16 is used. This variable is measured on a five-point scale: 1 is poor, 5 is excellent. This variable will be entered as a dummy-variable, with a separate dummy representing each category. Furthermore we control for marital status. We know from prior research that married people live longer, healthier lives than their non-married counterparts. According to Waite and Gallagher (2002) this effect is quite the same for men and women. One of the reasons is that a marriage encourages people to lower their level of risky behaviours (Waite & Gallagher, 2002). Marital status is entered as a dummy variable with two categories: married or non-married. The category married consists of married heads of the family as well as permanently cohabitants. Non-married heads are either divorced, widowed or single. 61% of the sample is married, 39% non-married. Also education level of father and mother are entered as a control variable in the regression. Education level of parents is obviously an important explanation of the educational attainment of their children (e.g. Davis-Kean, 2005), but it cannot be affected by it. Education level of parents is entered as a dummy-variable with two categories. The first category runs from zero to twelve grades (which indicates a high-school diploma), The second category indicates a grade upward of the 12^{th}. 21% of the fathers and 19% of the mothers have more than 12 grades of education completed.

To measure dietary awareness, we use eight variables regarding people's awareness of diet-related health problems and recommended dietary guidelines. People are asked if they are aware of the "food guide pyramid", the "5-a-day programme", and "dietary guidelines for Americans". They are also asked if they are aware of health problems caused by being overweight, by eating too much fat, by not eating enough calcium, not enough fibre, or too much cholesterol. Answers are 'yes' or 'no'. Descriptive statistics can be found in table 1. Cognitive dissonance can be an important caveat here: individuals may differ in the extent to which they report they know about what is harmful to them as a function of their behaviours. In the case of smoking, studies show different results. Sometimes people even overestimate the health risks (Cutle & Lleras-Muney, 2010). Most important here however is that it is not known whether these biases differ by education.

Table 1
Descriptive Statistics Dietary Awareness

Awareness of:	Answer 'yes' (%)	Answer 'no' (%)
Dietary guidelines for Americans	31	69
5-a-day programme	27	73
Food guide pyramid	55	45
Health problems caused by eating too much fat	85	15
Health problems caused by not eating enough fibre	54	46
Health problems caused by not eating enough calcium	81	19
Health problems caused by eating too much cholesterol	84	16
Health problems caused by being overweight	87	13

To measure the effect of the additional explanatory variables, as mentioned above, on the present state of health the variable 'present state of health of the head' will be used as health measure. This variable will be taken from the years 1999 and 2001, as will be explained in the methodology section. Only heads are taken into account, because they report their health themselves. The heads report the health of their wives. Health state of wives is therefore only indirectly reported and will not be taken into account for reasons of reliability. Heads are almost all man, therefore we don't correct for gender in this analysis. Present state of health is measured on a five-point scale: 1 is excellent, 5 is poor. For methodological reasons we define 1 and 2 as 'good' and 3, 4 and 5 as 'bad'. Because this variable is self-reported, it is difficult to determine the weight of the categories themselves and the differences between them. Approximately 43% of the people report their health as being bad, 57% as being good.

5. Methodology

This paper will test the effect of dietary awareness on the education-health gradient in diet related health behaviour. In testing this we first analyse the effect of dietary awareness on different health behaviours regarding diet. The additional explanatory variables (BMI, people's importance of choosing a diet with plenty of fruits and vegetables, the importance of choosing a diet low in saturated fat, the importance of choosing a diet with adequate fibre, the importance of a diet with plenty of grain products and the importance to maintain a healthy weight) will be used as dependent variable. We first estimate the effect of education on the different diet-related health behaviours:

$$HB_i = \beta_0 + \beta_1 * Education + \gamma Z_i + \varepsilon_i,$$

where HB_i are the measures of health-behaviour. β_1 measures the effect of education on health behaviour. Z_i are the control variables as mentioned in the data section and ε_i is the error term. Although most of the measures of health behaviour are binary, we will use a linear regression model, because the percentage change in coefficients will be analysed here. These coefficients can now be interpreted and compared directly. Using instead a binary probit model to estimate the effect of education on the binary measures of health behaviour will lead to roughly equivalent changes of the education-health gradient, see appendix table A.1.

Second, the knowledge questions Y_i will be added to the regression and then again the effect of the education variable will be measured:

$$HB_i = \beta_0 + \beta_1 * Education + \gamma Z_i + \delta Y_i + \varepsilon_i.$$

Table 2 shows the results of this regression analysis.

Where this first analysis measures the effect of dietary awareness on different health behaviours, we will now measure the effect of these health behaviours on the present state of health. If dietary awareness improves different health behaviours and these health behaviours have a significant impact on the present state of health, we are able to determine another part of the education-health gradient by measuring the influence of dietary awareness. The regression equation is as follows:

$$HS_{2001} = \beta_0 + \beta_1 * Education_i + \alpha X_i + \gamma Z_i + HS_{1999} + \varepsilon_i.$$

By doing this the explanatory power of the different health behaviours on the present state of health can be measured. Because all kinds of health behaviours are measured in the 1999th wave of the PSID, we use the present state of health of the 2001th wave. We expect problems of reversed

causality by analysing health behaviour and present state of health in the same year. Health behaviour can have an impact on the present state of health in the same year, but it is not hard to imagine that the effect can also be exactly reversed: people adapt their health behaviour to their present state of health. In this case, the present state of health in the same year is an important explanation of health behaviour and we are not able to determine cause and effect. The present state of health in 2001 can instead never have an impact on reported health behaviours in 1999. If this variable will be used as independent variable and we control for the state of health in 1999, the additional effect of the health behaviour in 1999 on the present state of health can be measured.

The dependent variable HS_{2001} is the present state of health in 2001. The education-health gradient is given by β_1 and is entered as independent variable. It measures the effect of education on the state of health in 2001. Subsequently, the variables regarding different health behaviours, X_i, will be added to the regression one by one and the effect of a particular health behaviour variable on the present state of health will be measured. HS_{1999} is the present state of health in 1999, Z_i are the control variables previously mentioned and ε_i is an error term.

Table 3 shows the results of this second regression analysis.

6. Results

Table 2

The impact of dietary awareness on health behaviours

	Regression coefficients without knowledge questions		Regression coefficients with knowledge questions		
Dependent variable	Grades of education (β)		Grades of education (β)		Reduction
BMI	-0.175	**	-0.170	**	-3%
Importance fruits and veggies	0.014	**	0.007	**	-50%
Importance low-fat diet	0.011	**	0.004		-64%
Importance fibre	0.015	**	0.007	**	-53%
Importance grain products	0.012	**	0.007	*	-42%
Importance healthy weight	0.012	**	0.008	**	-33%

Control variables: Age, family income, ethnicity, state of health until age 16, marital status, education level of father and education level of mother.

* Indicates statistical signifance at the 10% level.

** Indicates statistical signifance at the 5% level.

Table 3

Effect of health behaviours 1999 on present state of health 2001

	Coefficient health behaviour variable (α)		Coefficient grades of education (β)	
Health behaviour variable				
BMI	-0.007	**	0.013	**
Importance fruits and veggies	0.025	**	0.013	**
Importance low-fat diet	-0.016		0.014	**
Importance fiber	0.025	**	0.013	**
Importance grain products	0.021	*	0.014	**
Importance healthy weight	0.028	**	0.014	**

Dependent variable: Present state of health 2001

Control variables: Present state of health 1999, age, family income, ethnicity, state of health until age 16, marital status, education level father and education level mother.

* Indicates statistical signifance at the 10% level.

** Indicates statistical signifance at the 5% level.

7. Interpretation

According to the theoretical models of Grossman and van Kippersluis & Galama, we expect the higher educated, with a higher level of human capital, to be healthier and be more engaged in healthy consumption. When analyzing the relationship between education and health, we focus on specific knowledge to explain diet related health behaviour.

At first, the impact of dietary awareness on different aspects of health behaviour has been measured. The results can be found in table 2. When estimating BMI we see that the number of grades of education completed has a significant negative effect. Every year of education is associated with a 0.175 reduction of BMI. This effect is roughly comparable to the effect Cutler and Lleras-Muney (2010) found. If the knowledge questions are added to the regression, the effect of grades of education on BMI reduces with 3%. Every year of education is now associated with a 0.170 reduction of BMI. This is a much smaller reduction of the education coefficient than we observe later on with the variables regarding the importance of a healthy diet. There a reduction of 30% to 64% can be observed. These latter variables are however binary and more directly linked to the knowledge questions regarding a healthy diet and health problems. In explaining a person's BMI there may be a lot of other factors than specific health knowledge.

We subsequently measure the effect of educational attainment and dietary knowledge on the variables regarding the importance of a healthy diet. In contrast to variable BMI, which is categorical, these variables are binary. We estimate if people do find it important, or not. As can be observed from table 2, the number of grades of education completed has a significant and positive effect on the answer that people find it important to eat enough fruits and veggies, to choose a diet low in saturated fat, to eat enough fibre, to eat enough grain products and to maintain a healthy weight. The effect of educational attainment seems to be still small, but these variables run from 0 to 1, so the effect of education should not be underestimated. Every year of education leads to a 0.011 to 0.015 increase in the possibility that people find it important to choose a healthy diet. The question is however what the underlying causes are of this relationship. Is specific knowledge part of the explanation? According to our model we observe a significant and large reduction of the effect of education on the possibility that people find it important to choose a healthy diet. The effect of educational attainment reduces to roughly 0.007/0.008. Especially when measuring if people find it important to choose a diet with plenty fruits and veggies, enough fibre, enough grain products and to maintain a healthy weight, we see that specific knowledge regarding dietary awareness and health problems seems to be 33% to 53% of the explanation. If we add the knowledge questions to measure the effect on the importance of people choosing a diet low in saturated fat, the education

coefficient turns out to be insignificant. The knowledge questions regarding diet and health problems here outweigh the effect of education.

Table 3 presents the results of the regression analysis regarding the effect of the different health behaviours on the present state of health. The first column mentions the different health behaviours. The second column mentions the effect of this particular health behaviour variable on the present state of health in 2001. The education-health gradient, the effect of education on the present state of health, is given in column three.

BMI in 1999 has a significant negative impact on the present state of health in 2001. Every extra point of BMI in 1999 leads to a 0.007 reduction of the possibility that people report themselves as being in good health. So after controlling for the current state of health in BMI has an important impact on the future state of health.

The effect of the reported importance of people in 1999 to choose a diet with plenty fruits and veggies, enough fibre, enough grain products and to maintain a healthy weight on the state of health in 2001 is roughly comparable. If people find this important, this leads to a significant increase of 0.021 to 0.028 in the possibility that they report themselves as being in good health two years later. This is the additional effect of the 1999th health behaviour on the present state of health in 2001, controlling for the 1999th present state of health.

The variable importance of choosing a diet low in saturated fat turns out to has an insignificant impact on the present state of health in 2001. Controlling for the present state of health in 1999, this variable has no additional effect of the present state of health in 2001.

Comparing the results of table 2 and 3 we can conclude that dietary awareness is a significant part of the effect of education on different health behaviours. The effect of educational attainment on BMI, the importance of choosing a diet with plenty of fruits and veggies, the importance of eating enough fibre, eating enough grain products and the importance to maintain a healthy weight is reduced by controlling for dietary awareness. From table 3 we can conclude that these health behaviours have a significant effect on the present state of health.

8. Conclusion

By using the PSID dataset we examine the research question: *To what extent is the education-health gradient in diet related health behaviour influenced by dietary awareness?*

The concept of 'allocative efficiency' according to the theory of Grossman, assumes that the higher educated produce more health because they choose a better mix of inputs. Van Kippersluis and Galama (2013) extend this model and divide consumption into two parts: healthy and unhealthy consumption. The depreciation rate of health capital, or the biological aging rate, decreases with healthy consumption, but increases with unhealthy consumption. In other words, the health cost increases in the 'unhealthiness' of a good. In this research we test to what extent specific knowledge of this (un)healthy consumption has influence on health behaviour and differs across educational groups.

Education-health gradients in diet related health behaviour in this research are significant and quite large. Controlling for age, family income, ethnicity, state of health until age 16, marital status, education level of father and education level of mother, higher educated have a lower BMI and find it more important to eat plenty of fruits and veggies, choose a diet low in saturated fat, eat enough fibre, eat enough grain products and to maintain a healthy weight. When estimating BMI we see that the number of grades of education completed has a significant negative effect. Every year of education is associated with a 0.175 reduction of BMI. This effect is roughly comparable to the effect Cutler and Lleras-Muney (2010) found.

To examine whether and to what extent this effect is influenced by specific knowledge, we add the awareness of people regarding certain dietary guidelines and diet related health problems to the regression. This specific knowledge turns out to be a large part of the explanation of the effect of education on different health behaviours. The effect of education on BMI, the importance of eating plenty of fruits and veggies, enough fibre, enough grain products, and to maintain a healthy weight is significantly reduced. From this we can conclude that dietary awareness is an important explanation of the relationship between educational attainment and diet related health behaviour.

Furthermore the effect of these diet related health behaviours on the present state of health are examined in order to determine whether and to what extent these variables have an influence on the present state of health. BMI, the importance of eating plenty of fruits and veggies, eating enough fibre, enough grain products, and to maintain a healthy weight, all measured in 1999 and controlled for the present state of health in 1999, have a significant effect on the present state of

health in 2001. These variables are therefore important in determining the education-health gradient in diet related health behaviour.

Combining the two empirical analyses we are able to conclude that dietary awareness has a large influence on diet related health behaviours and that these diet related health behaviours have a significant impact on the present state of health. So the education-health gradient in diet related health behaviour is to a quite large extent influenced by dietary awareness.

Limitations and recommendations

In this research intervening control variables are entered in an OLS regression to estimate their effect on the variable of interest: number of grades of education completed. Ravesteijn (2013) shows in his research however that the estimators of direct causal effects and of partial associations conditional on intervening variables are biased in cases where the intervening control variable is correlated with the error term. An instrumental variable approach could be used to overcome this possible bias. Because this research is an extension of the existing research we also use the same methods for reasons of comparability and interpretation of the results. Although Cutler and Lleras-Muney (2010) state that they explain a percentage share of the education-health gradient, we do not draw that exact conclusions, but show more general tendencies.

The PSID dataset which is used in this research has several limitations. At first, a lot of variables are self-reported and therefore quite subjective. The variables present state of health and importance of diet related health behaviour are self-reported and measured on a five-point scale. How people rate their behaviour and state of health may differ. Furthermore, objective measures could be included here, such as: 'how many times did you visit you general practitioner last year?' or 'how many times a week do you eat fruits or veggies?'.

Questions regarding diet related health behaviour and dietary awareness were only asked in 1999. A more extended panel study was therefore not possible. If diet related health behaviour and dietary awareness would be measured over several years, more conclusions can be drawn from comparing this to other (health) behaviours and their state of health over time.

In the 1999th wave of the PSID people were asked about their awareness of diet-related health problems and recommended dietary guidelines. We do not know the extent to which these answers reflect the depth of people's knowledge and beliefs. It can be the case that people may know the correct answer or only 'have heard of it' without knowing it exactly or believing in it strongly. More

in depth questions could be asked here to get a better measure of people's knowledge and believes regarding diet-related health problems and recommended dietary guidelines.

Dietary awareness seems to be a large part of the education-health gradient in diet related health behaviour. From a governmental perspective, this is a support for awareness campaigns to improve people's health behaviour and to reduce obesity and other diet related health problems.

Bibliography

Albouy, V., & Lequin, L. (2009). Does compulsory education lower mortality? *Journal of Health Economics, 28*(1), 155-168.

Arendt, J. N. (2005). Does education cause better health? A panel data analysis using school reforms for identification. *Economics of Education Review, 24*(2), 149-160.

Behrman, J. R., & Rosenzweig, M. R. (2004). Returns to Birthweight. *The Review of Economics and Statistics, 86*(2), 586-604.

Case, A., Fertig, A., & Paxson, C. (2005). The lasting impact of childhood health and circumstances . *Journal of Health Economics, 24*(2), 365-389.

CBS. (2008). *Gezondheid en zorg en cijfers 2008.* Den Haag/Heerlen: Centraal Bureau voor de Statistiek.

Census Bureau, U. S. (1999). *Educational Attainment in the United States 1999*. Downloaded june 20, 2013, van http://www.census.gov/hhes/socdemo/education/data/cps/1999/tables.html

Clark, D., & Royer, H. (2010). The Effect of Education on Adult Health and Mortality: Evidence from Britain. *NBER Working Paper, 16013*.

Conti, G., Heckman, J., & Urzua, S. (2010). The Education-Health Gradient. *The American Economic Review, 100*(2), 234-238.

Cutler, D. M. (2006). Education and Health: Evaluation Theories and Evidence. *NBER Working Paper, 12352*.

Cutler, D. M., & Lleras-Muney, A. (2010). Understanding differences in health behaviours by education. *Journal of Health Economics, 29*(1), 1-28.

Davis-Kean, P. E. (2005). The Influence of Parent Education and Family Income on Child Achievement: The Indirect Role of Parental Expectations and the Home Environment. *Journal of Family Psychology, 19*(2), 294-304.

Edwards, R. D. (2010). Health, Income, and the Timing of Education among Military Retirees. *NBER Working Paper, 15778*.

Eide, E. R., & Showalter, M. (2011). Estimating the relation between health and education: What do we know and what do we need to know? *Economics of Education Review, 30*(5), 778-791.

Fuchs, V. R. (1980). Time Preference and Health: an exploratory study. *NBER Working Paper, 539*.

Grossman, M. (1972). On the Concept of Health Capital and the Demand for Health. *Jounal of Political Economy, 80*(2), 223-255.

Grossman, M. (2000). *Handbook of Health Economics, Chapter 7.* Cambridge: North-Holland.

Kenkel, D. S. (1991). Health Behaviour, Health Knowledge and Schooling. *Journal of Political Economy, 99*(2), 287-305.

Kippersluis, H. v., & Galama, T. (2013). Why the Rich Drink More but Smoke Less: The Impact of Wealth on Health Behaviors. *Discussion Paper Tinbergen Institute, 35*, 1-53.

Kippersluis, H. v., O'Donell, O., & Doorslaer, E. v. (2011). Long-Run Returns to Education: Does Schooling Lead to an Extended Old Age? *The Journal of Human Resources, 46*(4), 695-721.

Lillard, D. R., & Eamon, M. (2010). Live and Learn or Learn and Live: Does education lead to longer lives? *Unpublished*.

Lleras-Muney, A. (2005). The Relationship Between Education and Adult Mortality in the United States. *The Review of Economic Studies, 72*(1), 189-221.

McGonagle, K. A., Schoeni, R. F., Sastry, N., & Freedman, V. A. (2012). The Panel Study of Income Dynamics: overview, recent innovations, and potential for life course research. *Longitudinal and Life Course Studies, 3*(2), 268-284.

Mokdad, A. H. (2004). Actual Causes of Death in the United States, 2000. *Journal of the American Medical Association, 291*(10), 289-298.

PSID. (2013). *Data on Food and Nutrition.* Michigan: PSID.Org.

PSID. (2013). *General Guide.* Michigan: PSID.Org.

Ravesteijn, B. (2013). Good, Bad, and Ugly Controls: When should we use control variables? *Unpubl.*

Waite, L., & Gallagher, M. (2002). *The Case for Marriage: Why Married People are Happier, Healthier and Better Off Financially.* New York: Random House Digital Inc.

Appendix A

Table A.1

The impact of dietary awareness on health behaviours measyred by Binary Probit Estimates

	Regression coefficients without knowledge questions		Regression coefficients with knowledge questions		
Dependent variable	Grades of education (β)		Grades of education (β)		Reduction
Importance fruits and veggies	0.037	*	0.020	*	-46%
Importance low-fat diet	0.028	*	0.011		-61%
Importance fibre	0.039	*	0.019	*	-51%
Importance grain products	0.033	*	0.019	*	-42%
Importance healthy weight	0.032	*	0.022	*	-31%

Control variables: Age, family income, ethnicity, state of health until age 16, marital status, education level of father and education level of mother.

* Indicates statistical signifance at the 5% level.