Unto Häkkinen, Marjo-Riitta Järvelin, Gunnar Rosenqvist and Jaana Laitinen

Health, Schooling and Lifestyles: New Evidence from the Northern Finland Birth Cohort

Centre for Health Economics at STAKES CHESS

CHESS, the Centre for Health Economics at STAKES, was founded in September 2002. CHESS undertakes high-quality health economics research into health-policy relevant issues. CHESS aims to promote interaction between researchers in Finland and abroad and to increase the impact of health economics research.

CHESS focuses on three research areas: 1) quality and efficiency of health services, 2) quality and efficiency of elderly care services, and 3) financing and provision of health services.

The CHESS Discussion Papers intends to make research available quickly to the academic community, health policy makers and health professionals both in Finland and internationally. Material published in Discussion Papers ranges from work in progress to papers submitted for journal publication. Comments and suggestions on Discussion Papers are encouraged and should be sent directly to the authors.

The authors

Unto Häkkinen is a research professor in CHESS, email: unto.hakkinen@stakes.fi

Marjo-Riitta Järvelin is a professor in the Department of Public Health and General Practice, University of Oulu and Department of Epidemiogy and Public Health, Imperial College London, email: marjo-riitta.jarvelin@oulu.fi

Gunnar Rosenqvist is a senior statistical adviser in CHESS and professor in Swedish School of Economics, Business Administration, email: gunnar.rosenqvist@stakes.fi

Jaana Laitinen is a docent and senior researcher in Oulu Regional Institute of Occupational Health, email: jaana.laitinen@ttl.fi

Summary

This was a longitudinal, general population study based on a Northern Finland 1966 Birth Cohort, using a structural equation approach to estimate the health production function and health input functions for four lifestyle variables (smoking, alcohol consumption, exercise and unhealthy diet) for males and females. In particular, we examined the productive and allocative effects of education on health. We used 15D, a generic measure of health-related quality of life, as a single index score measure but we also estimated models for some of its dimensions. Among the males, the important factors impacting on health were education and all the four lifestyle factors, as well as some exogenous variables at 31 years and variables describing parents' background, and health and behaviour at 14 years. An increase of five years in schooling increased the health score by 0.008, of which about 50% was due to direct effect and 50% due to indirect effects. Among the females, education does not impact on health, but health was affected by the use of alcohol, exercise and diet, but not by smoking.

Our results indicate that policy options that increase education among men will increase their health indirectly via healthier lifestyles. However, since the total effect was rather modest and the direct effect insignificant, an increase of schooling is not a cost-effective way to increase health.

Keywords. Health; education; lifestyle; longitudinal study; health production function

Contents

Introduction

Information on health determinants is one of the most important starting points for health policy. Various studies by economists and epidemiologists have tried to understand the relationship between health, schooling and other policy relevant factors. Most economic studies on health determinants are based on the estimation of reduced form equations, often using cross-sectional and rather crude health variables. In our study, a structural equation model of health determinants was developed using a unique longitudinal birth cohort study in order to estimate the relative effect of factors impacting health. Of special interest was modelling the relationship between health and schooling whilst taking into account lifestyle mediators. A verified positive causal link between schooling and health would, depending on its nature, imply the possibility of increasing the aggregate level of health either by increasing schooling or by increasing health education and other activities designed to encourage health habits.

Most economic studies on the topic are made in developing countries or in the USA [1], in which education and schooling systems differ from those in Europe. The Finnish system is closest to those in the other Nordic welfare-state countries in which socio-economic equity has been emphasised as a target for both the educational system and health system. In addition, in Finland, the participation of women in the labour force is high compared to many other countries, which also may affect the relationship between health and schooling. In the mid 1990s the educational level of Finns aged 25–35 was clearly higher than EU average and among females the educational level was one of the highest in the EU [2].

One concern in previous studies was related to measuring health status. Usually it is measured by indicators such as self-rated health [3] [4] [5], activity limitations [6] [7], restricted activity days [7] [6] and blood pressure [6]. We measured health by 15D. It is a measure for health-related quality of life (HRQOL) [8] [9, 10], which combines information on different dimensions of health into a single score. In addition, we estimated the effects of education and lifestyle variables on the dimensions of the 15D. In many respects, especially in terms of discriminating power (sensitivity) the properties of the 15D have been to be found superior to generally used profile and single index score instruments [11] [12] [10]. The 15D has recently been used as a standard to validate different methods concerning the problems associated with use of self-rated health measures [13]. The 15D has been and is used in many projects for evaluating health technology and is included in population surveys. Thus, the results of this study can be compared with those from these previous studies.

Theoretical framework

The economic literature describing health determinants follows on predominantly from Grosmann's 1972 contribution [14]. In this framework, the individual is seen as combining market and non-market inputs to yield an output of good health. The individual is assumed to choose a health lifestyle based on health effects and direct utility effects, subject to income and time constraints. The individual also determines his or her health, in part, through health lifestyle choices. Different theoretical [5] [4] models lead to a general model on determinants of health in a period t:

$$H_t = H(H_{t-1}, L, E, X) \hspace{3cm} (1)$$

expansion of schooling [3] [1] . where H_{t-1} is health status in t-1, E is education, L is lifestyle and X is the vector of exogenous characteristics.

When estimating the health production function, the effect of schooling is important from a policy perspective. If there is a high correlation between health and schooling, an increase in expenditure on education may be a cost-effective technique for increasing the aggregate level of health. The effect of schooling on health has been subject to a large amount of economic research [1] [15] [16]. It is common to distinguish the productive (direct) from the allocative effects of education on health. Productive efficiency refers to the fact that education leads to a larger health output from a given set of health input. The notion of allocative efficiency[a] suggests that a more educated person is likely to select more efficient inputs (such as lifestyles) to produce health. For example, schooling increases information about the true effects of health inputs. The more educated may have more knowledge about the harmful effects of cigarette smoking or about what constitutes an appropriate, healthy diet. The distinction between the two forms of efficiency is important for resource allocation: evidence in support of allocative efficiency will justify efforts encouraging healthy habits whereas evidence in support of productive efficiency will justify an

On the other hand, a positive correlation between health and schooling may be due to one or more unobservable variables such as genetics, personal factors or rates of time preference affecting both health and schooling in the same direction. Finally, it can be due to reverse causality, arguing that better health results in more schooling. In econometric terminology, Grossman[15] points out that both explanations can be seen as falling under the general rubric of biases due to unobserved heterogeneity among individuals. In the case of unobserved variables or reverse causality, the policy relevant effects of an increase in education are not valid.

In this study we will evaluate directly the productive and allocative efficiency effects and try to take into account possible reverse causality, as well as control for a possible unobserved common source.

Methodical questions

From a methodological point of view it should be noted that the health production function is a structural equation system, since health inputs may also be endogenous. Ordinary least squares (OLS) estimates of the parameters of the production function may be biased and inconsistent because the inputs are likely to be correlated with disturbance terms. Early research in this area assumed that reduced form equations could be estimated by OLS. Later research has questioned this procedure; in particular, that schooling is uncorrelated with the disturbance term for health in the reduced form [1]. The usual method is to first estimate the reduced form equation for health inputs and then, in the second stage, the input demand functions are substituted into the health production function. As shown by Rosenzweig and Schultz [18], such a two-stage procedure can also take into account omitted variables (population heterogeneity), assuming that variables used to predict inputs are not correlated with the error terms of the input equation or the production function. In the two-stage least squares models, there have been difficulties in calculating the predicted values of the endogenous inputs: Most instrument variables used in the first stage have turned out to be poor predictors and the second-stage results have been sensitive to the specific specifications employed [19] [20] [7].

We estimate all equations of the structural model simultaneously. This is done by the LISREL program [21], which provides the possibility to include, for example, latent variables, measurement errors in dependent and independent variables, correlation between measurement errors, simultaneity, and detailed effect decomposition. Estimation is done with maximum likelihood under a normality assumption. This approach allows direct testing of the endogeneity of inputs and makes it possible to calculate direct and indirect (i.e. the productive and allocative efficiency) effects, which are not possible to separate from each other in reduced form equations. The statistical tests and diagnostics included in the output of the program (e.g. modification indices) help the investigators to choose the specification.

In this study, by applying the LISREL approach to longitudinal data it was also possible to take into account possible reverse causality, since we had information on health status and education at adolescence [20] [15]. The third variable hypothesis is tested by allowing disturbances of health and education to correlate. The previous studies on the effects of controlling the unobserved heterogeneity are not clear. For example, in the US study, this third variable bias was not significant and results were inconsistent with the time preference hypothesis [6]. On the other hand, Gillesekie and Harrison (1998) reported that controlling for unobserved heterogeneity using a discrete factor random effects estimator has a substantial impact on parameter estimates [4]. At least this underlines the importance of careful model specification, including the selection of the relevant explanatory variables.

Data and variables

The data are based on a Northern Finland 1966 Birth Cohort study (http:/kelo.oulu.fi/NFBC). All births in the provinces of Oulu and Lapland in Northern Finland 1966 (96.3% of all 1966 births) were eligible (n=12 058 live births). The data include questionnaires, hospital records and other information collected from other registers [22, 23].

Data on parents' socio-demographic background factors were collected by questionnaire during the 24th to 28th gestational weeks. Data on the course of the pregnancy were prospectively recorded in the maternity records, and transferred by midwives onto study forms, as were data on birth and the newborn at the time of delivery. Data were also collected at 1 year from child welfare centres and at 14 years by adolescent questionnaires. The latter include questions concerning growth and health, living habits, school performance and family conditions.

The latest follow-up in 1998, at age 31, consisted of questionnaires to all offspring (77% response) and further examinations for those living in the original target area or in the area of the capital Helsinki when additional inquiries on health and quality of life were distributed. For the rest of the cohort population living in other parts of Finland, the same data (15D) were collected by mailed questionnaire. The data are described in the Appendix. The data used here included 1989 males and 2354 females.

Table 1 show the variables included in the final models. Health status was measured by an index score of 15 dimensions: mobility, vision, hearing, breathing, sleeping, eating, speech, elimination, usual activities, mental function, discomfort and symptoms, depression, distress, vitality, and sexual activity [8] [9, 10]. The valuation system of the 15D is based on an application of the multi-attribute utility theory. A set of utility or preference weights, elicited from the general public through a valuation procedure is used in an additive aggregation formula to generate the 15D score (a single index number) over all the dimensions.

Table 1. Description of variables and their means among males and females

	Males	Females
Endogenous variables (at 31 years of age)		
Health, 15D score (H)	0.962	0.950
Schooling, number of years of schooling (E)	12.2	12.5
Smoking, ordinal variable describing smoking habits	0.70	0.47
(= no smoking, 1= occasional smoking, 2= regular daily smoking)		
Alcohol, consumption of alcohol (grams) per day	13.3	5.2
Exercise, number of minutes of heavy training in a month	334	287
Diet, ordinal variable describing dietary habits	2.39	1.72
(0= healthy diet, 5= unhealthy features in diet)		
Health at birth and parents background variables (X)		
Birth weight ,1000 g	3,6	3.5
Mothers schooling, number of years of schooling	6.8	6.8
Fathers socio-economic class 1 at 14 years old,	0.14	0.13
dummy variable = 1 if socio-economic class 1		
Fathers socio-economic class 2 at 14 yeas old,	0.20	0.19
dummy variable = 1 if socio-economic class 2		
Father living in the family at 14 years old,	0.90	0.88
dummy variable = 1 if father living in the family		
Living in rural area,	0.68	0.66
dummy variable = 1 if living for rural area at time of birth		
Health and behaviour at 14 years old (Z)		
Smoking at 14 years old,	0.05	0.06
dummy variable = 1 if smoking at least onece a week		
Alcohol drinking at 14 years old,	0.02	0.03
dummy variable = 1 if drinking at least once in a month		
Exercise at 14 years old, number of sport activities in a month	14.0	9.28
Average grade in all subjects at school at 14 years old	7.46	8.04
Repeated years at school at 14 years old	0.02	0.01
Occurrence of mild illness of long duration	0.14	0.14
Occurrence of severe illness of long duration	0.09	0.10
Number of Illness days during the year at 14 years old	1.56	1.59
Exogenous variables at 31 years old (Y)		
Unemployment, dummy variable = 1 if unemployed	0.09	0.10
Total years of unemployment	0.54	0.52
Student, dummy variable = 1 if student	0.02	0.04
Number of children in family	1.06	1.38
Number of adults in family	1.83	1.80

The maximum index score is 1 (no problems on any dimensions) and the minimum score is 0 (being dead). The 15D score is defined as

$$v_H = \Sigma_j \, I_{jk}(x_{jk}) w_{jk}(x_{jk}) = \Sigma_j \, D_{jk}(x_{jk}) \qquad\qquad (2)$$

where $I_{jk}(x_{jk})$ is the average relative importance people attach to level k (k=1, .., 5) of dimension j (j=1, .., 15), and $w_{jk}(x_{jk})$ is the average value people place on level k of dimension j. The main analysis is made using the 15D score as the dependent variable. Additional analyses were also made using the scores of individual dimensions as a dependent endogenous variable (Figure 1).

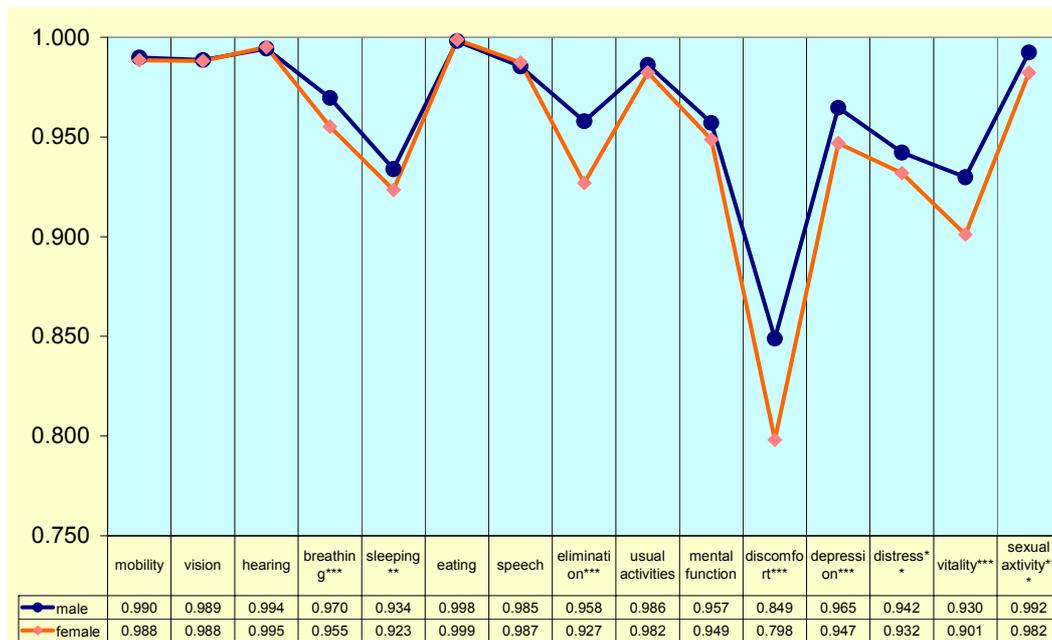|	mobility	vision	hearing	breathing***	sleeping**	eating	speech	elimination***	usual activities	mental function	discomfort***	depression***	distress**	vitality***	sexual axtivity***
male	0.990	0.989	0.994	0.970	0.934	0.998	0.985	0.958	0.986	0.957	0.849	0.965	0.942	0.930	0.992
female	0.988	0.988	0.995	0.955	0.923	0.999	0.987	0.927	0.982	0.949	0.798	0.947	0.932	0.901	0.982

Figure 1. The 15D profiles of the OULU Cohort 1966 population at 31 years old. Mean scores of each dimension among males and females. The scores are standardised so that the highest level of each dimension has a value of 1. Asterisks indicates statistical differences in mean scores between the genders (*=p<0.05, **= p<0.01, ***= p<0.001)

Lifestyle variables (diet, alcohol consumption, exercise, and smoking) as well as other background variables were ascertained at the 31-year follow-up as a part of the larger postal questionnaire sent to all cohort members. Data on food consumption was gathered with a method commonly used in Finnish population surveys [24] [25].

The subjects were asked to consider their food consumption during the previous 6 months and to choose a suitable alternative on a structured 6-point scale. Data on the frequency of consumption of food rich in fibre (such as rye bread, fresh vegetables and salads, berries or fruit) and food rich in high saturated fats (such as sausages) were obtained. From this information, an ordinal six-class variable was constructed (0 = unhealthy diet, 5= healthy features of diet) [26]. For the diet variable that is observed on an ordinal scale we use the LISREL approach of assuming an underlying latent continuous variable that is normally distributed with a zero mean and a standard deviation of one [27].

The questions on alcohol measured the average frequency of consumption of beer, wine, and spirits during the last year, and the usual amount of alcohol consumed on one occasion. The amount of alcohol (grams) consumed per day (continuous variable) was calculated using the average estimates of alcohol content in beer, light wines, wines and spirits[26]. The frequency of smoking (number of cigarettes per day) and exercise (number of minutes of training) were calculated in a similar way using rather detailed questions. Exercise was also treated as a continuous variable. Since distribution of smoking was rather skewed with a large number of zeros it was treated as an ordinal variable including three values (0= no smoking, 1= occasional smoking, 2= regular daily smoking)

Education was measured by the years of schooling prior to the 31-year follow-up, which were calculated from the education register data linked to cohort data using the unique personal ID-number.

As can be seen from the Appendix, the study used data from about 36% of the original sample and about 37% of the cases who were alive in 1997. The 15D variable was available for more than 50% of the cases. Attrition for different reasons decreased the sample considerably. An analysis of the sample selection indicated that persons with lower education had a much higher probability to be excluded from our sample than persons with a higher education (Appendix).

Model specification

In this study, our analytical focus is on the health determinants of 31-year-olds. It is assumed that their independent rational behaviour started after the age of 14. Thus, many variables related to health, e.g. health-related behaviour as well as family background measured at the age of 14 years are predetermined (exogenous) in our model.

The empirical model building process proceeded in stages. First, the input function for each lifestyle variable was estimated separately. In addition, a separate function was estimated for education in order to evaluate the possible causal effects of health determinants through education. Finally, the health production function (1) was estimated. With longitudinal data, the timing of events constitutes a natural restriction on the direction of causal relationships – cause must precede effect. Hence, we can specify a system of equations which is recursive at least if we disregard possible covariance between error terms.

In summary our model consists of the following equations,

$$H = a_1 + a_2E + a_3DIET + a_4EXERCISE + a_5ALCOHOL + a_6SMOKING$$
$$+ a_7X_1 + a_8Z_1 + a_9Y_1 + \varepsilon_1 \tag{3}$$

$$E = b_1 + \quad b_3X_2 + \quad b_4Z_2 + b_5Y_2 \quad + \varepsilon_2 \tag{4}$$

$$DIET = c_1 + c_2E + c_3X_3 + c_4Z_3 + c_5Y_3 + \varepsilon_3 \tag{5}$$

$$EXERCISE = d_1 + d_2E + d_3X_4 + d_4Z_4 + d_5Y_4 + \varepsilon_4 \tag{6}$$

$$ALCOHOL = e_1 + e_2E + e_3X_5 + e_4Z_5 + e_5Y_5 + \varepsilon_5 \tag{7}$$

$$SMOKING = f_1 + f_2E + f_3X_6 + f_4Z_6 + f_5Y_6 + \varepsilon_6 \tag{8}$$

where H is health status as measured with 15D; E is education; DIET, EXERCISE, ALCOHOL and SMOKING are lifestyle variables; X, Y and Z are vectors of exogenous variables with X describing parents background and health at birth, Z health and behaviour at the age of 14 and Y exogenous variables at the age of 31, while ε_j are error terms.

The X, Y and Z vectors need to be specified. Neither the theoretical health production model nor the findings of other relevant studies give us complete guidance for each of the model equations on the exact choice of specific variables from the set available. We perform a general-to-specific specification search [28] with the aim of finding a model which fits the data well and in which the parameters are significant and substantially meaningful. Parameters with small t-values are eliminated and parameters with large modification indices are added [21]. In addition to the variables given in Table 1, for example, a number of variables describing parent's behaviour and family circumstances at time of birth were excluded since they were not significant and did not affect the coefficients of other variables.[b]

Before the estimation of the final models we made some preliminary analysis using a single equation (OLS) production function for health (where all variables in Table 1 were treated as exogenous) in order to get some guidance for the model specification. We analysed the functional form of the endogenous variable and the differences between the sexes of effects that lifestyle variables and exogenous variables had on health. For example, in some studies the effect of alcohol consumption on health has been found to be non-linear so that moderate drinking has had favourable health effects compared to non-drinking or heavy drinking. In our data we did not find evidence of non-linearity.

There were gender differences in mean values of endogenous and exogenous variables (Table 1). The total score of 15D as well as eight of its dimensions (Figure 1) were statistically higher among males compared to females while the opposite was found with respect to schooling. The preliminary analysis using a single equation (OLS) production function indicated significant sex differences (Chow test). A dummy variable test indicated that the differences between the males and females were related in particular to smoking (more negative effect among males), alcohol consumption (more negative effect among females) and being a student (more positive effect among females). Thus, models were estimated separately for males and females. We specified similar models for both genders in order make their comparison easy. In other words, if a variable was significant for one of the sexes, it was kept in the model for both of them.

The general structure of the model is shown in Figure 2. There is no clear theoretical basis to model the relationship between the lifestyle variables. Thus we end up with a specification in which we allowed the disturbances of lifestyle variables to correlate
This means that the system of equations is in fact block recursive. The LISREL approach allows us to control for possible unobserved latent variables by allowing the error terms of endogenous variables to be correlated with each other and by specifying specific factors. We tested the third variable hypotheses by allowing the error terms of the health and education equations (3) and (4), i.e. ε_1 and ε_2, to be correlated. However, for both genders, the covariation between these error terms was not statistically significant (t=0.65 among males and t=0.99 among females)**.** Thus there is no significant covariation left to be explained by a latent "third variable".
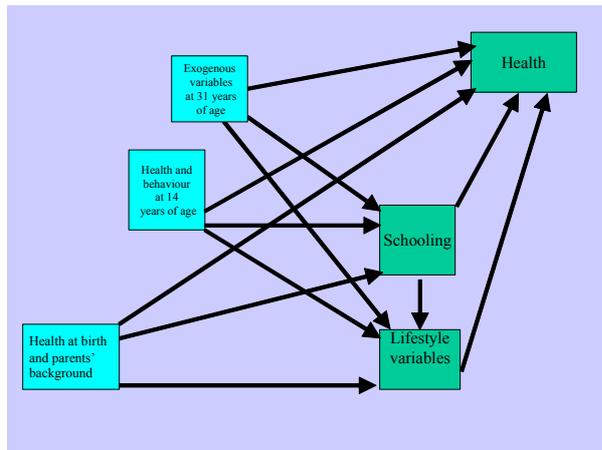
Figure 2. The structure of the model

In addition to looking at the direct and indirect effects of education on health, it could be of policy interest to distinguish between the total (cumulative) and dynamic (short-term) effects of education on health. Here, dynamic effects refer to the effect of education on the change of health status between two periods. As we have no measurements between the ages of 14 and 31, we cannot extract short-term effects, rather the estimated effects have to be regarded as cumulative ones. However, as pointed out by van Doorslaer [20], possible unobserved factors that have executed their whole effect on health status already at the age of 14 do not cause an omitted variable bias when estimating the health production function at 31, controlling for health status at the age of 14.

Results

The estimation results are displayed in Table 2 for males, in Table 3 for females and the total effects of exogenous variables on health are in Table 4. The goodness-of-fit of the models measured by the usual chi-square statistics as well as other measures were satisfactory.

TABLE 2. Estimation results, males (N=1989)

	Smoking	Alcohol	Exercise	Diet	Schooling	Health
Endogenous variables (at 31 years of age)						
Smoking						-0.002 *
Alcohol						-0.0002 ***
Exercise						0.00001 ***
Diet						-0.003 **
Schooling	-0.08 ***	-0.77 ***	9.3 *	-0.088 ***		0.0009
Health						
Exogenous variables						
Health at birth and parents background						
Birth weight				-0.08 *	-0.005	0.002
Mothers schooling					0.08 ***	
Fathers socio-economic class 1		2.7 *			0.6 ***.	
Fathers socio-economic class 2	0.03	2.7 **			0.2 *	
Father living in the family	-0.16 *					0.2
Living in rural area	-0.17 ***	-2.4 **		-0.09 *	-0.3 **	
Health and behaviour at 14 years of age						
Smoking at 14 years old	0.67 ***	5.4 *				
Alcohol at 14 years old	0.14	11.2 **				
Exercise at 14 years old			58 ***	-0.008 ***		0.0002 *
Average grade in all subjects at school	-0.19 ***			-0.14 ***	1.4 ***	
Repeated years at school	-0.06				-0.57 *	
Occurrence of mild illness of long duration					-0.06	-0.007 *
Occurrence of severe illness of long duration	0.09				-0.1	-0.012 ***
Number of illness days during the year	0.02 *			0.027 **	-0.02	-0.0007
Exogenous variables at 31 years of age						
Unemployment	-0.03	5.6 ***		118 ***		-0.01 **
Total years of unemployment	0.006 **	0.4				-0.0008
Student	0.20			0.38 **		-0.03 ***
Number of children	0.05 **	-1.2 ***	-31 ***	0.04 *	0.07 **	
Number of adults	-0.15 ***	-3.3 ***	6.4			
R^2	0.16	0.06	0.05	0.11	0.46	0.08

Chi-Square = 52.9 (P = 0.73)
Degrees of Freedom = 60
Root Mean Square Error of Approximation (RMSEA) = 0.09
Comparative Fit Index (CFI) = 1.000
Adjusted Goodness of Fit Index (AGFI) = 0.989
*=p<0.05, **= p<0.01, ***= p<0.001

TABLE 3. Estimation results, females (N=2354)

	Smoking	Alcohol	Exercise	Diet	Schooling	Health
Endogenous variables (at 31 years of age)						
Smoking						0.0001
Alcohol						-0.0009 ***
Exercise						0.000006 *
Diet						-0.005 ***
Schooling	-0.07 ***	-0.41 ***	-2.9	-0.047 ***		-0.00002
Health						
Exogenous variables						
Health at birth and parents background						
Birth weight				-0.05	-0.03	-0.004
Mothers schooling					0.08 ***	
Fathers socio-economic class 1		2.4 ***			0.4 ***.	
Fathers socio-economic class 2	0.10 *	-0.14			0.1	
Father living in the family	-0.27 ***				0.3 **	
Living in rural area	-0.13 ***	-0.10		-0.14 ***	-0.12	
Health and behaviour at 14 years of age						
Smoking at 14 years old	0.71 ***	1.6 *				
Alcohol at 14 years old	0.24 *	1.9				
Exercise at 14 years old			0.42 ***	-0.05 *		0.00007
Average grade in all subjects at school	-0.29 ***			-0.11 ***	1.26 ***	
Repeated years at school	-0.65 ***				-0.36	
Occurrence of mild illness of long duration					-0.06	-0.008 **
Occurrence of severe illness of long duration	-0.13 *				0.05	-0.009 **
Number of illness days during the year	0.02 *			-0.007	-0.03 *	-0.001 ***
Exogenous variables at 31 years of age						
Unemployment	0.16 *	-0.27	74 ***			-0.0008
Total years of unemployment	-0.04 *	-0.51 ***				-0.002 **
Student	0.2 *			0.28 *		-0.005
Number of children	-0.089 ***	-1.39 ***	-43 **	-0.06 ***	-0.20 ***	
Number of adults	-0.20 ***	-1.13 ***	26 ***			
R²	0.23	0.08	0.04	0.04	0.43	0.06

Chi-Square = 68.3 (P = 0.22)
Degrees of Freedom = 60
Root Mean Square Error of Approximation (RMSEA) = 0.008
Comparative Fit Index (CFI) = 0.998
Adjusted Goodness of Fit Index (AGFI) = 0.987
*=p<0.05, **= p<0.01, ***= p<0.001

Among the males, use of alcohol (-), smoking (-), (unhealthy) diet (-), unemployment (-), being a student (-), indices for the occurrence of long-duration illnesses (-) and exercise both at the ages of 31 and 14 years were directly and statistically significantly related to health (Table 2). We find a significant total effect on health for the variables of previous childhood: mothers' schooling (+); father living in a family (+); smoking (-); drinking habits (-); increased average school grades (+). In addition, the total effect of number of adults (+) as well as children (-) in family at 31 years were also significant (Table 4). Schooling was positively related to health, but the relationship was not statistically significant (P=0.07) . There was a clear indication of the allocative effects of schooling, since schooling was related to the lifestyle variable in a health promoting way.

Table 4. Total effects of exogenous variables on health

	Males	Females
Parents background		
Birth weight	0.002	-0.003
Mothers schooling	0.0001 **	0.00004
Fathers socio-economic class 1	0.0003	-0.0019 **
Fathers socio-economic class 2	-0.0003	0.0003
Father living in the family	0.0007 *	-0.0003
Living at rural area	0.0008	0.0006
Health and behaviour at 14 years old		
Smoking at 14 years old	-0.003 **	-0.0003
Alcohol drinking at 14 years old	-0.003 *	-0.001
Exercise at 14 years old	0.0003 **	0.0001
Average grade in all subjects at school	0.003 ***	0.0007
Repeated years at school at 14 years old	-0.0008	-0.001
Occurrence of mild illness of long duration	-0.007 *	-0.008 **
Occurrence of severe illness of long duration	-0.012 ***	-0.009 **
Number of illness days during the year	-0.0009	-0.0013 **
Exogenous variables at 31 years old		
Unemployment	-0.01 **	0.0002
Total years of unemployment	-0.0001	-0.002 *
Student	-0.033 ***	-0.006
Number of children in family	-0.0004 *	0.001 ***
Number of adults in family	0.001 ***	0.0004

As was expected from our preliminary analysis there were clear differences in the results between the genders. The most important difference is that smoking and schooling were not associated with health among females as they were among males. On the other hand, alcohol consumption, exercise and diet were related to health in a similar way among females as among males, but the negative effects of alcohol on health was much greater among females. Among females, education was related in a health promoting way to smoking, alcohol consumption and diet, as among men, but not to exercise. In addition, the mechanism and effects associated with parents' socioeconomic status, behaviour at 14 years, being a student, unemployment and family structure were different. For example, among females, the total effect of the fathers' good socio-economic status on health was negative since this variable was associated with increasing alcohol consumption, and although it was positively related with education, as noted, education was not significantly related with health among females. Finally, it is worth a mention that the number of children had a positive effect only among females whereas the opposite held for males. The positive effect of the number of adults in the family was significant only among males.

In Table 5 we have illustrated the results by calculating the total effects as well as the direct and indirect effects of schooling on the 15D score. In addition we have also illustrated practically the effects of the lifestyle variables. For the ordinal variables (smoking and diet) the effects have been calculated on the basis of normal scores in the form of class means of the assumed normal variable [27]. The indirect effects of schooling reflect the effects of schooling on health via lifestyle variables.

Table 5. Total direct and indirect effect of education and lifestyle variables on health

	Males			Females		
	Total	Direct	Indirect	Total	Direct	Indirect
Schooling increase in five years of schooling	0.008 ***	0.004	0.004 ***	0.002	0.000	0.002
Smoking change of smoking habits from no smoking to regular daily smoking	-0.004 *	-0.004 *	-	0.003	0.003	-
Alcohol consumption increase in consumption by 16 gr.(one bottle of beer) per day	-0.004 ***	-0.004 ***	-	-0.014 *** *	-0.014 ***	-
Exercise increase in training by one hour per week	0.003 ***	0.003 ***	-	0.001 ***	0.001 *	-
Diet change of diet from healthy to unhealthy (from score 0-1 to score 4-5)	-0.009 **	-0.009 **	-	-0.015	-0.015 ***	-

*=p<0.05, **= p<0.01, ***= p<0.001

Among males, an increase of five years of schooling increases the health score by 0.008 (i.e. increases health by about 1%), of which about half is due to direct and half to indirect effects. Among the females, the small total effect is due to indirect effect. Among the males, the total effect of change of diet from unhealthy to healthy had about the same effect on health as an increase in schooling of five years. Even greater health effects can be obtained among the females by a similar change in diet or by decreasing alcohol consumption by 16 grams (about 1–1.5 bottle of beer) per day.

In general, although most changes are statistically significant, their practical importance is not high. A change of 0.02–0.03 in the score has been observed to be such that people can feel the difference [10]. We found such effects only in variables describing student status (among males) and alcohol consumption (change of over 24 grams per day among females).

In order to get a more detailed picture of the relationship between health, schooling and lifestyle, the models developed earlier were applied to some of the dimensions of the 15D. Since our study population consists of young adults whose health status was considerably good, the analysis is only sensible for those dimensions where there is sufficient variation. For example, in dimensions such as mobility, hearing, vision, and eating, the mean score was rather near 1 (Figure 1), with most of the individual values concentrated at the highest level. Table 6 describes the estimation results of those dimensions for which we got significant effects for schooling and lifestyle variables. In most cases, the same model developed for the overall index score also fitted the data, in terms of chi-square statistics, for the dimensions evaluated. Only in the models for sleeping (both genders), as well as breathing, elimination, mental functioning and depression (among females) did we make modest modifications to the original model in order to get a satisfactory goodness-of-fit.

According to dimension specific analysis among males, the total effects of schooling were positive and significant in breathing, usual activities, mental functioning, and

discomfort and symptoms. Only in elimination did we find any significant negative effect of schooling on health. In usual activities and mental functioning, the positive total effect of schooling was due to a direct effect, but in breathing, discomfort and symptoms there was also a significant indirect effect. In sleeping, eliminating, depression, distress and vitality the indirect effect of schooling was positive and significant, but not so great that it resulted in a significant total effect. Among the dimensions the mean score was lowest in discomfort and symptoms and was under 0.95 also in sleeping distress and vitality, i.e. in dimensions where indirect effects are very significant (Figure 1). Thus we can assume that any positive indirect effects of schooling on the total 15D score are mainly due to these effects.

Among males, the negative effects of smoking on health (total 15D) seem to be due to negative effects in four dimensions (breathing, sleeping usual activities, and vitality), negative effects of alcohol consumption in six (breathing, sleeping, elimination, depression, distress, and vitality) and unhealthy diet in four dimensions (sleeping, discomfort and symptoms, depression, and vitality), whereas the positive effect of exercise was due to effects found in seven dimensions (breathing, elimination, mental functioning, discomforts and symptoms, depression, distress and vitality).

Among the females, the total effect of schooling was positive only for mental functioning, which was also the only dimension in which we found an indication of productive (direct) effect. The overall positive indirect effect of schooling on the total 15D can be explained by the corresponding indirect effect of breathing, sleeping, discomfort and symptoms, depression, distress and vitality.

Among females the observed overall rather strong negative effect of alcohol consumption on health seems to be explained by negative effects of alcohol in eight dimensions (breathing, sleeping, usual activities, mental functioning, discomfort and symptoms, depression, distress and vitality). Exercise has a positive effect in two (mental functioning, and vitality) and unhealthy diet has a negative effect in six dimensions (breathing, sleeping, discomfort and symptoms, depression, distress and vitality). Among the females, smoking was not related to the total score (15D). We find a negative significant effect of smoking only in breathing. However, rather unexpected results are the positive effects of smoking to usual activities and mental functioning.

Table 6. Statistically significant effects of education and lifestyle variables on dimensions of 15d

	Males			Females		
	Total	Direct	Indirect	Total	Direct	Indirect
Breathing						
Schooling	+++	+++	+++			+++
Smoking	- -	- -		- -	- -	
Alcohol	- - -	- - -		-	-	
Exercise	++	++				
Diet				- - -	- - -	
Sleeping						
Schooling			+++			++
Smoking	-	-				
Alcohol	- - -	- - -		---	---	
Diet	- - -	- - -		---	---	
Elimination						
Schooling	- -	--	+			
Alcohol	-	-				
Exercise	+	+				
Usual activities						
Schooling	++	++				
Smoking	-	-		+	+	
Alcohol				---	---	
Exercise						
Mental functioning						
Schooling	++	++		+	+	
Smoking				++	++	
Alcohol				---	---	
Exercise	+	+		+	+	
Discomfort and symptoms						
Schooling	+	+	++			+
Alcohol				- - -	- - -	
Exercise	+	+				
Diet	--	--		- -	- -	
Depression						
Schooling			+++			+++
Alcohol	---	---		- - -	- - -	
Exercise	+	+				
Diet	-	-		- - -	- - -	
Distress						
Schooling			+++			+++
Alcohol	- - -	- - -		- - -	- - -	
Exercise	+	+				
Diet				- - -	- - -	
Vitality						
Schooling			+++			+++
Smoking	-	-				
Alcohol	-	-		- - -	- - -	
Exercise	+++	+++		+	+	
Diet	- - -	- - -		- - -	- - -	

+ positive and significant effect (p<0.05), ++ positive and significant effect(p<0.01),+++ positive and significant effect (p<0.001),- negative and significant effect (p<0.05), --negative and significant effect(p<0.01),---negative and significant effect (p<0.001)

Conclusions

In this study, we have analysed the relationship between schooling and health using longitudinal data with a generic measure of health-related quality of life. As in earlier economic studies [4, 6, 7], our results confirm significant effects of education on health, but only among males.

But even among males, the effect was rather modest. Among females, whose educational level was higher, we could find no significant total effects of schooling on health.

The richness of the data, the character of the health variable and the methodological approach allow us to distinguish between the productive and allocative effects of education, which has only been done before—as far as we know—in one US study [4]. Our results concerning the importance of the allocative effect among men disagree with the results of the US study, where the productive effects were clearly greater than the allocative effects. Using our approach, it was also possible to describe in more detail the mechanism underlying the causal relationships, as well as to find out the most important dimensions of health on which education and lifestyle had effects.

Theoretically, our health measure (15D) defines literally billions of mutually exclusive health states, thus having a great potential for discriminatory power and responsiveness to small changes. We find many statistically significant although not always clinically important effects. The latter can be due to the fact that the analysis is based on a sample that is relatively healthy, indicating that effects are rather small compared to the effects of some health care interventions [10]. On the other hand, results can be seen also as an indication that 15D and some of its dimensions are very sensitive even in a relatively healthy population.

Although the data and measures of health status were better than those available in many other studies, there are still factors that should be taken into account when considering the results. Firstly, as usual in a longitudinal study with a long follow-up, the number of missing cases was significant. The analysis of non-responses indicated that persons with a lower level of education clearly have a greater possibility of being excluded from the sample. It can be assumed that these persons also have lower health status than those included. One could argue that our results possibly underestimate the effect of schooling on health.

Secondly, the fact that the study includes a cohort from Northern Finland may create some caveats in generalising the results. For example, cultural and religious factors in Northern Finland differ from those of the rest of the country, which may affect the role and effects of lifestyle variables in health production. However, according to a recent study among Finnish adolescents, the regional differences in health behaviour reflect more socio-economic variation than cultural factors [29]. For example, adolescents' smoking and physical activity were determined merely by individual characteristics, and of those behaviours (drinking and diet) that where influenced by socio-economic context

(i.e. cultural factors), the relationship was more moderate than that between individual characteristics and health behaviour [30]. In addition, a considerable part of the cohort has moved from Northern Finland, which also reduces the effects of cultural factors.

Thirdly, our health production function does not include health care utilisation, which may result in some biased effects of schooling and health behaviour on health. However, health care utilisation is rather modest among young adults. According to Finnish studies (e.g [31]), the use of health care among children as well as young adults is not significantly related to education. Thus, we can assume that the effect of omitting the use of health care from the health production function is not very important for studying the relation between health and education.

Fourthly, concerning the estimated effects of lifestyle variables on health one should take into account that their negative health effects will appear at later stages of life. This might be a reason that we do not find much negative effects of smoking on health. The reverse might hold for the use of alcohol since the positive effects of moderate drinking have been indicated to reduce cardiovascular disease, which usually emerges in later stages of life. However, in this study, we find among both genders very clear and significant effects of schooling on diet, smoking and alcohol consumption. Since these habits usually begin early in life, it is possible that our results underestimate the allocative effects of schooling on health which will be realised in later stages of life.

Many epidemiological studies have demonstrated a strong, negative association between education and health status as measured by mortality or morbidity. For example, a Dutch longitudinal study of men found an inverse relation between education level and survival even after confounding effects of height and health score were taken into account [32]. When lifestyle variables were included in the analysis, the results suggest that the higher prevalence of major risk factors among those with a lower educational level is not the dominant mediating mechanism that can explain educational disparities in health status [33]. More generally, lifestyle variables usually explain a rather modest proportion of the socio-economic gradient in mortality or morbidity. Contoyaniss and Jones [5] indicated that the failure of epidemiological analyses to account for unobserved heterogeneity could explain their low estimates of the relevance of lifestyle in the relationship between socio-economic status and health. In our study, we found that among males the indirect effect of schooling was statistically significant and explained about 50% of the total effect. Among the females, we find only significant indirect effect.

Our results indicate that policy options that increase education among men will increase their health indirectly via healthier lifestyles. However, since the total effect was rather modest and the direct effect insignificant, an increase of schooling is not a cost-effective way to increase health. A better alternative will be paying more attention to activities promoting directly healthy lifestyles. Among females our results did not give any support to increasing the years schooling and indicates the importance of drinking behaviour as a target for health promotion.

The results of studies made in other countries have been interpreted to indicate causality between more schooling and better health [1] [16]. The young adults' and particularly female's internationally high educational status in Finland might be a reason why we find only a modest effect of schooling on health and the non-existence of such effects among females. However, since education still strongly affects health behaviour, one can argue

that a high education status as such does not guarantee the socioeconomic equity in health which has been the main targets of health policy during recent decades in Finland.

REFERENCES

1.   Grossman M, Kaestner R. *Effects of Education on Health. In:* Behrman JR, Stacey N, editors. The social benefits of education. Ann Arbor: University of Michigan press; 1997. p. 69-123.
2.   OECD. *Education at glance, Oecd indicators 1977.* Paris; 1997.
3.   Leigh PJ. Direct and direct effects of education on health. *Social Science& Medicine* 1983;**17**(4):227-234.
4.   Gillesekie DB, Harrison AL. The effects of endogenous health inputs on the relationship between health and education. *Economics of Education* 1998;**17**(3):279-297.
5.   Contoyannis P, Jones A. Socio-economic status, health and lifestyle. *Journal of Health Economics* 2004;**23**:965-995.
6.   Berger M, Leigh P. Schooling, Self-selection, and Health. *The Journal of Human Resources* 1988;**24**:343-455.
7.   Kenkel DS. Should you eat your breakfast? *Health Economics* 1995;**4**(1):15-29.
8.   Sintonen H. The 15D measure of health related quality of life : reliability, validity and sensitivity of its health state descriptive system. Melbourne: National Centre for Health Program evaluation.; 1994. Report No.: Working paper 42.
9.   Sintonen H. The 15D-measure of health related quality of life. II feasibility, reliability and validity of its valuation system. Melbourne: National Centre of for Health Program Evaluation; 1995. Report No.: Working Paper 42.
      10. Sintonen H. The 15D instrument of health-related quality of life: properties and applications. *Annals of Medicine* 2001;**33**:328-336.
11.   Stavem K. Reliability, validity and responsiveness of two multiattribute utility measures in patients with chronic obstructive pulmonary disease. *Qual Life Res* 1999;**8**:45.
12.   Hawthorne G, Richardson J, A. DN. A comparison of the Assessment of Quality of Life (AQoL) with four others generic utility instruments. *Annals of Medicine* 2001;**33**:358-370.
13.   Lauridsen J, Christiansen T, Häkkinen U. Measuring inequality in self-reported health - discussion of a recently suggested approach using Finnish data. *Health Economics* 2004;**13**(7):725-732.
14.   Grossman M. *The demand for health: a theoretical and empirical investigation.* New York: NBER; 1972.
15.   Grossman M. The human capital model. In: Culyer AJ, Newhouse JP, editors. *Handbook of Health Economics*, Volume 1A. Amsterdam: Elsevier; 2000. p. 348-408.
16.   Grossman M. The demand for health, 30 years later: a very personal retrospective and prospective reflection. *Journal of Health economics* 2004;**23**(4):629-636.
17.   Muurinen J-M. Demand for health: a generalised Grossman Model. *Journal of Health Economics* 1982;**1**:5-28.
18.   Rosenzweig MR, Schultz PT. The behaviour mothers as inputs to child health: the determinants of births weight , gestation, and rate of fetal growth. In: Fuchs V, editor. *Economic aspects of health.* Chicago: University Chicago Press; 1982. p. 53-92.
19.   Kemna H. Working conditions and the relationship between schooling and health. *Journal of Health Economics* 1987;**6**:189-210.
20.   van Doorslaer EKA. *Health, knowledge and the demand for medicalcare.* Assen: Van Corcum&Comp; 1987.
21.   Jöreskog K, Sörbom D. *Lisrel 8. User's reference guide.* Chicago: Scientific Software International, Inc; 1996.

22.   Rantakallio P. Groups at risk in low birth weight infants and prenatal mortality. *Acta Paediatr   Scand* 1969;**193**:1-71.
23.   Laitinen J, Power C, Jarvelin M. Family social class, maternal body mass index, childhood body mass index, and age at menarche as predictors of adult obesity. *Am J Clin Nutr* 2001;**74**:287-294.
24.   National Public Health Institute. *Finravinto* 1997 (the 1997 dietary survey of Finland). Helsinki; 1998.
25.   Helakorpi   S,   Uutela   A,   Puska   P.   *Suomalaisten   aikuisväestön terveyskäyttäytyminen, kevät 1997, Health behaviour among Finnish adult population,* Spring 1997. Helsinki: National Public Health Institute; 1997.
26.   Laitinen J, Pietilainen K, Wadsworth M, Sovio U, Jarvelin M. Predictors of abdominal obesity among 31-y-old men and women born in Northern Finland in 1966. *Eur J Clin Nutr* 2004;**58**(1):180-90.
27.   Jöreskog K, Sörbom D. *Prelis 2: User's Reference Guide*. Linconwood: Scientific Software International,; 1999.
28.   Hendry DF. *Dynamic Econometrics*. Oxford: Oxford University Press; 1995.
29.   Karvonen S. *The regional context of health behaviour among Finnish adoleslcents*. Jyväskylä: Stakes, Research Reports 76; 1997.
30.   Karvonen S, Rimpelä A. Socio-regional context as a determinant of adsolescents' health behaviour in Finland. *Social Sciences & Medicine* 1996;**43**(10):1467-1474.
31.   Häkkinen U. Change in determinants of use of physician services in Finland between 1987 and 1996. *Social Science & Medicine* 2002;**55**:1523-1537.
32.   Doornbos G, Kromhout D. Educational level and mortality in a 32-year follow-up study of 18-year old men in the Netherlands. *International Journal of Epidemiology* 1990;**19**(2):374-379.
33.   Lanz PM, Lynch JW, House JS, Lepkowski JM, Mero RP, Musick MA, et al. Socioeconomic disparities in the health change in a longitudinal study of US adults: the role of health-risk behaviours. *Social Science & Medicine* 2001;**56**(1):29-40.

Appendix: Description of data used and analysis of nonresponse rate.

Table A1

	Number of cases
Total original cohort	12058 (12231, all births)
Alive in 1997	11877
Postal questionnaire in 1997	
- Respondents to which the postal questionnaire was sent	11541
- Returned questionnaires	8764
Clinical examination 1997 (those living in provinces of Oulu, Lapland and Capital area of Finland	
- Invited	8463
- Participated	6066
- Cases with completed 15D	5606
Middle of Finland (other than Oulu, Lapland and Capital area**) postal questionnaire of 15D 1997**	
- Cases sent	2000
- Cases with completed 15D	1381
Total number of cases with completed 15D	6987
Total number of cases included in this study (after deletion of cases with missing data on all variables)	4343

Table A1 describes original data and the data used in this study. As usual in a long lasting panel study, attrition is significant. As can be seen from the Table there are many sources for nonresponse. Using data collected from the education register, it was possible to analyse the determinants of nonresponse. From the register we got information on 9699 (4844 males and 4855 females) original cohort persons of which we also had data collected at 14 years of age. Nonresponse was higher among the males (59%) than among the females (52%). A logit regression using parents' background variables as well as health and behaviour variables at 14 years (see Table 1) indicated that the probability of being included in our sample was, among males, strongly related to schooling (+) and also statistically significantly to occurrence of severe illness (-), smoking at 14 years old (-), average grade (+), repeated years of schooling (-) and living in a rural area (+). Among females, the probability was also related strongly to schooling (+), and statistically significantly also to average grade (+), repeated schooling (-), and to the variable describing whether the father lived in a family (+).

The nonresponse rate was studied also by logit regression whilst including both genders in the data, producing a significant dummy variable for gender (indicating higher nonresponse among males) . We also compared the logit models for attrition between the genders with a Chow type test, which indicated significant differencexs between the genders. A dummy variable test illustrated statistically significant differences in effects between the genders in four variables (smoking at 14 years, occurrence of severe illness, living at rural areas and fathers socioeconomic class being 2 at 14 years old).

NOTES:

---

[a] Muurinen refers to this concept as use-related deprecation [17].
[b] Based on referees suggestion we included birth weight into the models, although the variable was not statistically significant