Willing to be healthy? On the health effects of smoking, drinking and an unbalanced diet. A multivariate probit approach *

Udo Schneider and Brit S. Schneider

June 2009

Preliminary version! Please do not cite or quote!

Correspondence address:

University of Bayreuth
Department of Law and Economics
Institute of Public Finance
D-95440 Bayreuth
Germany
Phone: +49-921-552929
Fax: +49-921-555821
Email: Udo.Schneider@uni-bayreuth.de
Brit.Schneider@uni-bayreuth.de

* We would like to thank Andrew Jones, Nigel Rice and Volker Ulrich for helpful comments on this paper – of course, any remaining errors are entirely ours.

Abstract

In the theoretical literature, time and medical services are essential input factors for producing health capital. Based on the idea that individuals can influence their health through health investments, we use a multi-level approach to consider the role of health relevant behaviour in the health production process. We assume that besides medical services, health relevant behaviour and outcomes are central determinants of the assessment of health. The estimation of a multivariate probit model using German data confirms most of the formulated hypotheses about the influence of education, income and labour force status on health relevant behaviour and health. The results indicate that health policy should encourage investments in education which correspond to investments in health.

JEL-Classification: I12, C31, D12

Keywords: health behaviour; health assessment; multivariate probit; education; labour force participation

1 Introduction

Individual's health perception is an important factor for the demand for medical services, and is itself driven by health behaviour. Concerning different behaviour patterns, the economic costs related to e. g. smoking, lack of exercise and obesity are immense. The demand for medical care due to obesity is about 2-8 % of overall health care budgets in Europe (cf. WHO (2005)). For Germany, Sander and Bergemann (2003) estimate the total costs of obesity at € 2,709-5,682 million, including the direct costs of obesity and the indirect costs of four co-morbidities: non-insulin-dependent diabetes mellitus, myocardial infarction, hypertension, and stroke. Therefore, total costs must even be greater if overweight[1] is additionally taken into account.[2] To counteract rising health care expenditures, it is necessary to start a process of rethinking in order to achieve changes in attitudes towards health. Health policy generally tries to implement personal responsibility through financial incentives, for example using demand side cost-sharing rules. For these to be effective, knowledge about the determinants of health related behaviour is essential. It must be assured that a lack of patients' responsibility can be separated from missing abilities to handle the own health capital stock.

In the theoretical literature about health production, health can be viewed as part of the human capital stock. This argumentation follows the idea that a good health status increases humans' productivity which is necessary to produce goods and services. On the one hand, healthy workers can spend more time on further education and use their knowledge more effectively. On the other hand, illness leads to absence from work. Hence, it is of main interest how to enlarge individuals' health capital stock. Similar to Grossman (1970), we assume that the individual's state of health is a result of investing in health but also of health depreciation. In this context, health investment is often primarily seen as the consumption of curative medical services. Furthermore, prevention of illness and therefore healthy behaviour is essential for a long term good health status in general. Thus, it should also be taken into account. Following the Grossman approach, health underlies a depreciation rate which depends on age as well as on adverse health behaviours (cf. Zweifel & Breyer (1997)).

Besides medical services, education is fundamental for being healthy. First, this assumption follows the idea that education leads to a higher allocative efficiency, meaning the knowledge of which medical services to consume. Second, education goes along with a higher productive

[1] Obesity means a body mass index (BMI=weight in kg/(height in m)²) of 30 or greater. For being overweight, the BMI ranges between a value of 25 and 29.9 (cf. World Health Organization (2003) for a classification in more detail).
[2] The expenditures of the German sickness funds related to nutrition based illnesses are about 30 % of total treatment expenditures. Additionally, high costs arise due to respiratory illnesses and cardiovascular diseases.

efficiency, e. g. how to utilize pharmaceuticals correctly. Third, more educated individuals have higher opportunity costs of illness due to a higher loss of wages.

Moreover, a person's labour force status and income play a central role for the possibility to act in a healthful manner. As these are not elements in the Grossman health production function, a reformulation seems appropriate. Therefore, more input factors are considered in our framework. For the subsequent analysis, producing health is viewed as a multi-level process in order to distinguish between health behavioural aspects, health outcome and the individual assessment of health.

The paper is organized as follows: In the next chapter, the importance of health behaviour for health production is analysed. Here, the basic idea of the Grossman model is presented and extended with respect to education, working hours, labour force status and behavioural aspects. In addition, we formulate five testable hypotheses about the structure of the health production process. The data set and the multivariate probit estimation method are described in detail in chapter 3 and 4, respectively. The fifth section addresses the estimation results. Moreover, we explicitly refer to the hypotheses presented in chapter 2. The paper ends with a conclusion and some policy implications.

2 On the Relationship between Health Related Behaviour and Health

In his seminal work, Grossman (1970) deals with the individual's ability to affect its own health status. In order to gain healthy time, people use medical care and time as input factors in the health production function. The efficiency of this production is also determined through the individual's education level because education may help to produce health or to allocate medical services in a more efficient way. A better health increases the individual's utility by reducing illness. Moreover, it is required for gaining labour income, which is itself a requisite for purchasing medical care and other consumer goods. The main point of this framework is that besides education E, time TH and medical services M are the basic inputs in the production of health capital. The equation of Grossman's household production function for investments in health I can be written as

$$I = I(M, TH; E) \tag{2.1}$$

The resulting health stock increases with health investment but is reduced by depreciation δH, where δ represents the proportional depreciation rate and H the health capital stock:

$$\Delta H = I - \delta H$$

In the model mentioned above it is assumed that individuals choose the inputs in the health production function under certainty, so that there is no doubt about the quality of medical treatment or the depreciation rate of health capital. Furthermore, time and medical inputs are treated as perfect substitutes insofar that they are related to the same productivity. These assumptions are somewhat restrictive and therefore have been subject to criticism (cf. e. g. Muurinen (1982), Dardanoni & Wagstaff (1990), Selden (1993) and McGuire, Henderson & Mooney (1997)). Moreover, most individual behaviours have an impact on the health capital stock in general. This means that there are several other determinants not related to medical care which are essential for a good health status. For example, a balanced diet is a product of the consumed food as a whole, which is not a health input in the sense of Grossman. Sanitation is another example. Finally, individuals can either reduce or increase their own depreciation rate by doing sports, consuming alcohol or smoking.

These aspects lead to the question which factors predominately determine health relevant behaviour. In general, education yields to better health knowledge which is important to understand the health effects of one's actions. Better educated individuals e. g. know more about the long-term health risks of overweight, so it can be expected that they pay more attention to their nutrition to watch their weight. In addition, better educated people know about the hazardous consequences of smoking. Kenkel (1991) for instance shows that education has a significant negative impact on smoking and alcohol drinking, while the impact on doing sports is significantly positive.

Furthermore, labour force participation should be considered as another important factor on health relevant behaviour. First, long working hours reduce leisure time and health investment activities. There is less time disposable for recreation, doing sports or even consuming some health services for preventive purposes. Second, the kind of work a person does is decisive for its health depreciation rate (cf. Leigh (1983) and Kemna (1987)). On the one hand, people like blue collar workers who do physically exhausting jobs may be less willing or less able to do some sport after work. On the other hand, managers mostly do a stressful job with long working hours. As a consequence, in spite of higher education they may have a high probability to be a smoker or to have excess weight (cf. Schofield (1996), Shields (2000) for an overview). Therefore, working conditions and education both are central determinants of the depreciation rate. In addition, both labour time and education determine earned income, which is itself fundamental for health related behaviour. Low income individuals tend to consume cheaper meals with low nutritional value. As a consequence, the risk of overweight or even obesity is much higher at low incomes (cf. Bhattacharya, Currie & Haider (2004)). In addition to that, doing sports often requires some equipment, so that time and money are rather complements than substitutes in this aspect.

Apart from these three direct effects of labour force participation on health relevant behaviour the opportunity costs if illness rise with labour income, which means that illness reduces current and future earnings. Because of this, the benefits of healthful activities are largest for well educated people with high labour income (cf. Gilleskie & Harrison (1998), Schneider, Schneider & Ulrich (2007)). Unemployed don't have these opportunity costs, so there are no economic incentives for health investment activities. Mathers & Schofield (1998) show for instance that beside a poorer mental health, those who are unemployed have greater odds of suffering chronic illnesses. There is also some evidence that unemployed people tend to higher levels of smoking, alcohol use and poor diet.

One frequently observed result of omitted health investment is overweight or even obesity, which both may have dramatic consequences for an individual's health: The most severe diseases related to heavy body weight are "hypertension and hyperlipidaemia (major risk factors), coronary heart disease, ischaemic stroke, type 2 diabetes, certain types of cancer, osteoporosis and psychosocial problems" (World Health Organization WHO (2005), p. 1), among others. As a consequence, people who are overweight may lose productivity due to a poor health status and therefore labour income, while on the other hand, based on their insurance status they may face higher out of pocket health expenditures.

The findings above lead to a health production function which is different from Grossman's approach. Here, we distinguish between health behaviour, health outcome and self-assessed health. Health behaviour HB can be described through consumption patterns that influence health directly, like smoking, alcohol consumption, and nutrition. These indicators depend on the individual's labour force status W, educational variables E, and other personal characteristics V (e. g. income, risk taking behaviour):

$$HB = HB(W, E, V) \tag{2.2}$$

Second, health outcome (HO) is an objective measure of the health status, which is itself influenced through the health behaviour described above. In addition, education and again other individual characteristics X are of particular importance:

$$HO = (HB, E, X) \qquad (2.3)$$

Last, self-assessed health *SAH* measures the individual's perception of its health capital stock. It is a function of health outcome and health behaviour. Furthermore, the demand for medical care *M,* personal attributes *Z*, and education *E* are elements of this third health equation:[3]

$$SAH = SAH(HO, HB, M, Z, E) \qquad (2.4)$$

These three equations (2.2)-(2.4) form a recursive system where some explaining variables of the *HB*-equation only have indirect effects on health outcome and self-assessed health.[4] The structure of the full model is depicted in figure 1.

< figure 1 around here>

Here, the solid lines show the impact of the exogenous variables on the dependent variables. The influence of first health behaviour on health outcome and second of both variables on *SAH* is represented through dashed lines. Last, the consumption of medical services is only considered in the *SAH* equation due to endogeneity problems.

With regards to this theoretical framework we are able to formulate the following hypotheses, which will be analyzed empirically in section 5:

H1: Unemployment leads to adverse health activities, while being in the labour force has a positive impact on health relevant behaviour.

H2: Long working hours lead to adverse health behaviour.

H3: Health behaviour is essential for health outcome, and both factors explain individual's health assessment.

[3] Kenkel (1995) estimates the effect of lifestyles on adult health using indicators such as eating breakfast, smoking and exercise as inputs into the production of health as well as different output measures. In addition, he also investigates the role of schooling for the production of good health.

[4] Here, the set of variables in *Z* and *X*, can be describes as a subset of *V*, respectively. Generally, *Z* and *X* may also contain variables that are not included in *V*.

H4: Income has positive effects on health behaviour, health outcome and the individual assessment of health.

H5: Education is a central determinant of health relevant behaviour, health outcome and self assessed health.

3 Data

All data are from the German Socio-Economic Panel (SOEP), a representative longitudinal study of private households in Germany.[5] Explicitly, we use the 2006 wave where different variables concerning health status and health behaviour are included. Beginning with health relevant activities, smoking, alcohol consumption, and nutrition are incorporated in the data-set.[6] While the variable smoker is measured as a binary variable in the dataset, alcohol consumption and nutrition are both categorical variables. To treat all indicators equally and with respect to the estimation strategy, we transform the two categorical into binary indicators. In detail, the variable fast food measures whether an individual does not follow a health conscious diet at all. The variable alcohol takes the value 1 if the respondent drinks at least one of the following beverages regularly: beer, wine or champagne, spirits, and mixed drinks.[7] Moreover, smoker takes the value 1 if the respondent states that she is smoking.

Concerning the health outcome, the respondents are asked to give information about their body weight and height. We use this information to calculate the body-mass index as a measure of health outcome (cf. World Health Organization (2003)). Moreover, to account for the effects of overweight, we generate a binary indicator for the age-adjusted body-mass index that indicates overweight with respect to the individual's age (cf. National Research Council (1989)).

The self-assessed health variable in the dataset might be vulnerable to a reporting bias. For the correction of the self-assessed health, questions that rely on the so-called SF-12v2 indicators (cf. Andersen, Mühlbacher, Nübling, Schupp & Wagner (2007)) are used.[8] In the dataset, the original twelve questions are transformed into eight new subscales using either one or two of the original questions. We use seven of them to explain the self-assed health of the individuals

---

[5] The data used in this publication were made available to us by the German Socio-Economic Panel Study (SOEP) at the German Institute for Economic Research (DIW), Berlin.
[6] Similar indicators are used by Vita, Terry, Hubert & Fries (1998). They show that mortality rates and disability risks depend on tobacco consumption, physical exercises, and nutrition.
[7] The questionnaire of the SOEP gives no information about the quantitative meaning of regularly; the alternative answers were occasionally, seldom, and never.
[8] The SF-12v2 is a health related questionnaire especially on aspects of quality of life covering the dimensions physical and mental health (cf. Andersen et al. (2007)).

together with personal characteristics and health care utilization in order to overcome a possible reporting bias of self-assessed health. These results are used to compute a new health stock variable which takes the value 1 if health is assessed above average and 0 otherwise (for further information, see the appendix).

The independent variables can be divided into two different categories: The first category contains predisposing variables like gender and nationality. Four age categories are included to capture the deterioration of health with age due to comorbidity risks. In addition, partnership and children are indicators for the family status of the respondent. All these variables are binary ones. Second, socioeconomic variables are included to explain the economic environment. The first variables in this category determine the money disposable for consumer and health care goods, namely income, economic worries and unemployment. For the income position, we use the logarithm of the net household income to correct for the skewness of the density function. As the size of the household differs for the respondents, we correct for this effect using the household equivalent income. This income is computed by dividing the net household income through the square root of the household size (cf. German Council of Economic Experts (2000)). Furthermore, working conditions are implemented as explanatory variables. Precisely, working hours are used to explain the trade-off between work, health investment, and leisure, and to control for working conditions which are not covered through the variables ‘civil servant’ and ‘blue collar’, respectively. To control for the expected nonlinear effects, dummy variables for different classes of working hours are created. The remaining variables are educational levels and a variable representing the willingness to take risks (see table 1).

< table 1 around here>

If one takes a look at the descriptive statistics in table 2 below, it is obvious that there are differences with respect to health related behaviour. In detail, 50% do not eat a balanced diet, nearly 30% are smokers, and still 16% drink alcohol regularly. Almost 36% of the respondents are overweight, and 56.6% range their health above average. Without age-adjustment, 54.5% must be classified as overweight.

<table 2 around here>

Although the Socio-Economic Panel is a representative dataset in general, in our sample Eastern Germans are overrepresented (29 vs. 20 %) due to non-responses and drop outs. Concerning the insurance status, 10.8 % are fully privately insured which corresponds to the actual level in Germany. Taking into account that Eastern Germans are overrepresented in this dataset and that their earned income is below average, the share of fully private insured might be slightly biased upwards.

4 Estimation method

For the estimation of the described problem we use a simultaneous equations model. With respect to the structure of the theoretical model and the dependent variables, a recursive multivariate probit model is applied. This can be seen as a generalization of the bivariate probit model presented in Maddala (1983)). Generally, the multivariate probit model can be written as:

$$\begin{aligned} y^*_{1i} &= \beta_1' X_{1i} + \varepsilon_{1i} \\ y^*_{2i} &= \beta_2' X_{2i} + \varepsilon_{2i} \\ &\vdots \\ y^*_{Mi} &= \beta_M' X_{Mi} + \varepsilon_{Mi} \end{aligned} \qquad (4.1)$$

Here, we have m=1,…, M equations and i=1,.., N observations. For the latent dependent variables, we assume that

$$y_m = \begin{cases} 1 & \text{if} \quad y^*_m > 0 \\ 0 & \text{otherwise} \end{cases} \quad , m = 1 , \dots , M$$

Moreover, X_{mi} are vectors of exogenous variables, β_m the associated parameter vectors and $\varepsilon_{1i}, \ldots, \varepsilon_{Mi}$ are normally distributed errors with a constant variance var(ε_{mi})=1. As a result of the theoretical considerations about the health production process, we identify three classes of binary dependent variables: first, health behaviour of the individual, second, health outcome and third, a measure of self-assessed health. The recursive structure of the multivariate probit represents the distinction between the dependent variables as follows. The equations for the health behaviour variables are modelled as reduced-form equations. The health outcome equation is a structural equation with the health behaviour variables as explanatory factors. Last, in the self-assessed health equation health behaviour and health outcome are included as regressors. Therefore, we estimate the following system of three reduced-form and two structural equations (cf. Balia and Jones (2008):

$$\begin{aligned}
y_{1i}^* &= \beta_1' X_{1i} + \varepsilon_{1i} \\
y_{2i}^* &= \beta_2' X_{2i} + \varepsilon_{2i} \\
y_{3i}^* &= \beta_3' X_{3i} + \varepsilon_{3i} \\
y_{4i}^* &= \delta_{41}\, y_{1i} + \delta_{42}\, y_{2i} + \delta_{43}\, y_{3i} + \beta_4' X_{4i} + \varepsilon_{4i} \\
y_{5i}^* &= \delta_{51}\, y_{1i} + \delta_{52}\, y_{2i} + \delta_{53}\, y_{3i} + \delta_{54}\, y_{4i} + \beta_5' X_{5i} + \varepsilon_{5i}
\end{aligned} \tag{4.2}$$

The covariance between the error terms of equations *j* and *k* can be expressed as correlations $\rho_{jk}=\rho_{kj}$ that have to be estimated (cf. Cappellari & Jenkins (2003)). They measure in how far unobserved factors influence health relevant behaviour, health outcome and self-assessed health simultaneously. All equations in (4.2) can be estimated separately as single probit models but the estimated coefficients would be inefficient because the correlation between the error terms is neglected. Only in the case of independent error terms ε_{mi} (all ρ are not significantly different from zero) it is possible to deal with the above model as independent equations (cf. Maddala (1983)).[9]

The estimation of a recursive multivariate probit model requires further assumptions for the identification of the model parameters. For the model given in equation (4.2), Maddala (1983)

[9] Knapp & Seaks (1998) provide a Hausman test for the exogeneity of a dummy variable in a probit model, which is based on the estimated correlation coefficients.

shows that the number of parameters to be estimated is larger than the number of probabilities, even if constant terms are the only exogenous variables. In this case, the parameters in the structural equation are not identified. To answer this problem, Maddala proposes that at least one of the reduced-form exogenous variables must not be included in the structural equations as explanatory variables.[10]

Given the estimation at hand, we therefore impose exclusion restrictions. For the reduced form equations, we use the complete set of predisposing and socioeconomic variables. It is necessary to exclude at least one of these variables from the structural equations to identify the model correctly, which means that some variables must not be included in the equations for overweight and health. The same is true for the transition from the overweight to the health equation. It is assumed that working hours, unemployment status and other labour variables have only indirect effects on overweight and health through the individual's health relevant behaviour. Moreover, by excluding these factors we avoid endogeneity problems which might arise from health problems or the assessment of health. Similarly, current education and risk taking do not enter these equations.

Concerning the health status equation, it is supposed that current education is without any direct impact. Furthermore, we run three models with different restrictions for the health equation to identify the model parameters. First, we exclude the three variables living in a partnership, children and renovation. Second, we include the renovation and third, family variables are considered but renovation is omitted. Following equation (2.4), variables describing the demand for medical care are explanatory factors for the variable health. In detail, we only use health indicators in the correction for the reporting bias and refrain from including these variables in the multivariate probit estimation. The same is true for the variables describing the health insurance of the respondent (see the appendix).

For the selection of the appropriate set of exclusion restrictions measures of goodness-of-fit are used. First, the Akaike information criterion (AIC) and second, the Bayesian information criterion (BIC) are employed (cf. Long (1997)). The Akaike information criterion is based on the log-likelihood of the estimated model. It represents the trade-off between the goodness of the estimation, by means of the log-likelihood and the parsimony of the specification, which is given through the number of estimated parameters. The BIC has the weight on more parsimonious models than the AIC. These information criteria can be used to compare different

[10] On the contrary, the structural equations may contain variables not included in the reduced-form equations. According to Wilde (2000) the parameters of the model are identified as long as there is at least one varying exogenous regressor.

model specifications. That model which possesses the lowest value of the AIC or BIC is chosen as the best (cf. Verbeek (2008)).

The results for the information criteria for the three specifications are presented in table 3. For all estimated models, the measures of fit are relatively close together and point into the same direction. Moreover, the estimation results for the different models are comparable, implying only small differences in the parameter estimates and standard errors. The preferred model is specification III where besides the health insurance variable only renovation is excluded.

<table 3 around here>

5 Estimation results

The first three columns of table 4 show the partial effects for the reduced form behaviour equations estimated in the full recursive model, using the multivariate probit specification. For the smoker equation, most of the explaining variables are of significance, whereas the impact factors on nutrition are not that clearly to determine. Nevertheless, nearly all significant parameters of the smoker and fast food equations point in the same direction. In contrast, most of the alcohol effects are of opposite sign.

In detail, we observe a negative income effect on smoking, while drinking alcohol is positively influenced. Against our expectations, an impact on nutrition cannot be found. These results correspond to the estimated effects of the economic worries variable which is positive for smoking and fast food but negative for alcohol consumption. Hence, this indicator may be a comprehensive proxy for the socioeconomic status of the respondent.

As stated in hypothesis 1, being unemployed goes along with adverse behaviours. The same effect can be found for blue collar workers in the smoking and fast food equations. With respect to working hours, a significant positive impact on smoking and fast food consumption can be found (hypothesis 2). This goes along with the idea that long working hours cause stress and reduce recreation time.

For all three adverse health behaviours, we find strong effects of the education level. From theory, education is a necessity to understand the long-term consequences of unhealthy behaviour. This idea can be supported only for the smoking and fast food equations. Moreover and against our expectations, the probability of frequent drinking rises with educational attainment. This shows that alcohol drinking is socially accepted. For individuals currently in education, the expected effects also appear in the alcohol equation.

<table 4 around here>

The fourth column in table 4 presents the results for the structural health outcome equation. Starting with the endogenous variables, alcohol and fast food show a highly significant impact on the probability of being overweight. In contrast, smoking is without any significance. Although the overweight variable is already age-adjusted, getting older goes along with a positive effect. Hence, by adjusting the dependent variable an overestimation of the age effects is prevented. As expected, high income individuals face a lower risk of being overweight. This finding corresponds to the strong positive effect of economic worries. Higher education, i. e. a high school or university degree, goes along with a lower probability of a high BMI. For the insurance status, no weight effect can be found.

The health equation contains all of the endogenous variables. Here, only the variables smoker and fast food are of significance. In detail, smokers have a superior perception of their health status. One possible interpretation is that this group does not take adverse health consequences of their behaviour into account. In line with expectations, the effect of fast food is negative. Concerning the insignificant effect of alcohol, we do not distinguish between the different kinds of beverages in our sample. In addition, drinking beer or wine may go along with positive health effects (c. f. Mukamal, Chiuve & Rimm (2006)) or psychic well-being as part of the health status. Therefore, it may be that these positive health effects counteract possible negative effects of frequent spirit drinking. In addition, we expected that overweight goes along with a lot of health limitations and that the individuals are conscious of them when rating their health status. This cannot be found in our data. Given that negative consequences of overweight on health are well known, the non-significance of this indicator implies that individuals are not aware of their health limitations. Another explanation may be that they only face higher future health risks which are not incorporated in our SAH-variable. To sum up, hypothesis 3 can only partly be supported.

The age results in the health equation go along with the common finding that ageing leads to higher co-morbidity risks. Moreover, in the first three equations it is apparent that women act in a more healthy way while the coefficient is significantly negative in the health perception equation. This corresponds to the common empirical finding that women are more concerned about their health status and that they consult a physician more often.

Hypothesis 4 states positive income effects in all five equations. Our results support this hypothesis except for the alcohol and the fast food equation. In detail, the positive effect on drinking alcohol goes along with the education effects described above. Against our expectations, income is without direct impact in the fast food equation. Here, only the economic worries variable shows the expected sign. Therefore, not following a health-conscious diet is not a matter of the disposable income. Again following our expectations, persons with higher income have a lower probability of being overweight. Moreover, they are more likely to report their health better than average. In equation five, all education variables show a positive sign. Once more, except for the alcohol equation the postulated effects can be found. As a result, hypothesis 5 cannot be rejected.

Concerning the estimation technique, the main advantage of the multivariate probit model is that it considers a possible dependency between the equations. Therefore, it is possible to test whether health behaviour and health outcome are endogenous for health assessment. The five estimated equations involve ten correlation coefficients ρ_{jk} which measure the pairwise correlation between the three health relevant behaviour indicators, the health outcome and the health variable. Seven of the estimated ten correlation coefficients are of significance. This result implies that the equations are not stochastically independent and that single probit estimates would have led to inefficient standard errors. Moreover, the dependent variables of the first four equations are not exogenous in the health equation. The null hypothesis of no joint significance of these parameters is rejected using a likelihood ratio test.

6 Conclusion and policy implications

In the previous analysis, health production is viewed as a multi-level process in order to distinguish between health behavioural aspects, health outcome and the assessment of health. From a theoretical point of view, the importance of the individual's health behaviour for the health production process is beyond controversy. To test for the determinants of health behaviour, health outcome and self-assessed health, we apply a multivariate probit approach consisting of three reduced form and two structural equations. By using this procedure, it is possible to account for the endogeneity of smoking, alcohol consumption, nutrition behaviour, and overweight for health perception. Moreover, especially the relevance of the level of education, the individual's income situation, and labour force participation is of main interest.

Many theoretical and empirical studies support the idea that education is one of the driving factors for good health behaviour. Here, this result does not hold for all reduced-form equations. While the probability of bad nutrition and smoking falls with education that for frequent drinking rises. Taking the negative influence of education on overweight and the positive effect on health assessment additionally into account, especially less educated individuals have to be informed about long-term consequences of adverse health behaviour to strengthen their personal responsibility with respect to health. Furthermore, income is essential with respect to the probability of overweight and for the self-assessment of the health status. It is remarkable that alcohol consumption differs from the other lifestyle variables in that especially the effects of the income variables are reversed.

Last, we assumed that people in the labour force show better health behaviours, while those who are involuntarily unemployed exhibit adverse health activities. In fact, the negative impact of unemployment is evident in the data. The effects of working hours on health behaviour are mixed. For alcohol consumption, there is no difference between those working and those not working voluntarily but for smoking and fast food, a positive relation exists. These effects do not support theory of higher opportunity costs of being in the labour force. In contrast, smoking may be an efficient way to cope with stress or to have a friendly get-together with colleagues. Moreover, long working hours go along with time limitations leading to adverse eating habits.

To sum up our findings, socioeconomic variables like education, income and labour force status are central elements for health relevant behaviour and for the health production process. In addition, the decision about health behaviour, the observed outcome and the individual health assessment are dependent. As a result, health policy should go beyond the narrow scope of demand for medical care. Many health problems seem to be related to individual behaviour which is often not taken into account. The related financial burden on the health care system is immense. This raises the question of encouraging personal responsibility. In the last years, health policy has attempted to strengthen personal responsibility through financial incentives. These cost-sharing instruments may help to reduce the demand for health care and health expenditures but suffer from neglecting the link between patient characteristics and health relevant behaviour. Hence, our results point the importance of education, income and labour force participation for health behaviour and health perception. Rather than solely focussing on financial incentives, health policy should encourage investments in education that correspond to investments in health to decrease health expenditures related to adverse health behaviour. Getting people out of unemployment may be an instrument to reduce income risks and economic worries, but in detail, operational promotion of good health must be fostered.

References

Andersen, H. H., Mühlbacher, A., Nübling, M., Schupp, J. & Wagner, G. (2007), Computation of standard values for physical and mental health scale scores using the SOEP version of SF-12v2. *Schmollers Jahrbuch,* 127, 171-182.

Balia, S. & Jones, A. M. (2008), Mortality, Lifestyle and Socio-Economic Status. *Journal of Health Economics*, 27, 1-26.

Bhattacharya, J., Currie, J. & Haider, S. (2004), Poverty, Food Insecurity, and Nutritional Outcomes in Children and Adults. *Journal of Health Economics,* 23, 839-862.

Cai, L. & Kalb, G. (2006), Health Status and Labour Force Participation: Evidence from Australia. *Health Economics,* 15, 241-261.

Cappellari, L. & Jenkins, S. P. (2003), Multivariate Probit Regression using Simulated Maximum Likelihood. *The Stata Journal,* 3, 278-294.

Dardanoni, V. & Wagstaff, A. (1990), Uncertainty and the Demand for Medical Care. *Journal of Health Economics,* 9, 23-38.

Disney, R., Emmerson, C. & Wakefield, M. (2006), Ill Health and Retirement in Britain: A Panel Data-based Analysis. *Journal of Health Economics*, 25, 621-649.

German Council of Economic Experts (2000), *Chancen auf einen höheren Wachstumspfad* [Chances of a higher growth path], Stuttgart: Melzer-Poeschel.

Gilleskie, D. B. & Harrison, A. L. (1998), The Effect of Endogenous Health Inputs on the Relationship between Health and Education. *Economics of Education Review,* 17, 279-297.

Grossman, M. (1970), *The Demand for Health: A Theoretical and Empirical Investigation*, Columbia University: Dissertation.

Hagan, R., Jones, A. M. & Rice, N. (2008), Health Shocks and the Hazard Rate of Early Retirement in the ECHP. *Swiss Journal of Economics and Statistics*, 144, 323-335.

Hernández-Quevo, C., Jones, A.M. & Rice, N. (2005), *Reporting bias and heterogeneity in self-assessed health. Evidence from the British Household Panel Survey*, HEDG Working Paper No. 05/04. University of York, York.

Jones, A.M. (2007), *Panel data methods and applications to health economics*, HEDG Working Paper No. 07/18. University of York, York.

Kemna, H. J. M. I. (1987), Working Conditions and the Relationship between Schooling and Health. *Journal of Health Economics,* 6, 189-210.

Kenkel, D. S. (1991), Health Behavior, Health Knowledge, and Schooling. *Journal of Political Economy,* 99, 286-305.

Kenkel, D. S. (1995), Should You Eat Breakfast? Estimates from Health Production Functions. *Health Economics,* 4, 15-29.

Knapp, L. G. & Seaks, T. G. (1998), A Hausman Test for a Dummy Variable in Probit. *Applied Economics Letters,* 5, 321-323.

Leigh, J. P. (1983), Direct and Indirect Effects of Education on Health. *Social Science and Medicine,* 17, 227-234.

Long, J. S. (1997), *Regression Models for Categorical and Limited Dependent Variables*, Sage: Thousand Oaks et al.

Maddala, G. S. (1983), *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge et al: Cambridge University Press.

Mathers, C. D. & Schofield, D. J. (1998), The Health Consequences of Unemployment: the Evidende. *The Medical Journal of Australia,* 168, 178-182.

McGuire, A., Henderson, J. & Mooney, G. (1997), *The Economics of Health Care. An Introductory Text*, London et al.: Routledge.

Mukamal, K. J., Chiuve, S. E. & Rimm, E. B. (2006), Alcohol Consumption and Risk for Coronary Heart Disease in Men With Healthy Lifestyles. *Archives of Internal Medicine,* 166(19), 2145-2150.

Muurinen, J. (1982), Demand for Health. *Journal of Health Economics,* 1, 5-28.

National Research Council (1989), *Diet and Health. Implications for Reducing Chronic Disease Risk.*, Washington D.C.: National Academy Press.

Sander, B. & Bergemann, R. (2003), Economic Burden of Obesity and its Complications in Germany. *European Journal of Health Economics,* 4, 248-253.

Schneider, B. S., Schneider, U. & Ulrich, V. (2007), Health and the Decision to Invest in Education. *Journal of Economics and Statistics,* 227(5+6), 725-745

Schofield, D. (1996), *The Impact of Employment and Hours of Work on Health Status and Health Service Use*, Discussion Paper No. 11, NATSEM, University of Canberra, Canberra.

Selden, T. M. (1993), Uncertainty and Health Care Spending by the Poor: The Health Capital Model revisited. *Journal of Health Economics,* 12, 109-115.

Shields, M. (2000), Long Working Hours and Health. *Perspectives in Labour and Income,* 12, 49-56.

Verbeek, M. (2008), *A Guide to Modern Econometrics*, 3rd. edition, Chichester et al: Wiley.

Vita, A. J., Terry, R. B., Hubert, H. B. & Fries, J. (1998), Aging, Health Risks, and Cumulative Disability. *The New England Journal of Medicine,* 338, 1035-1041.

Wilde, J. (2000), Identification of multiple equation probit models with endogenous dummy regressors. *Economics Letters,* 69, 309-312.

World Health Organization (2003), *Diet, Nutrition and the Prevention of Chronic Diseases. Report of a Joint WHO/FAO Expert Consultation*, WHO Technical Report Series No. 916, Geneva.

World Health Organization WHO (2005), *The Challenge of Obesity in the WHO European Region*, Fact sheet EURO/13/05.

Zweifel, P. & Breyer, F. (1997), *Health Economics*, New York and Oxford: Oxford University Press.

Appendix

Modelling individual's health stock

Self-reported measures of health and their validity have caused a considerable debate (cf. Jones (2007)). Concerning the health variables, the self-assessed health variable might be vulnerable to a reporting bias because of anticipation and measurement errors (cf. Hagan, Jones & Rice (2008) and Hernández-Quevo, Jones & Rice (2005)). The original self-assessed health variable in the dataset (SAH) is a five-point scale variable ranging from very good to bad. To correct for a possible reporting bias, we apply a technique proposed by Disney et al. (2006). We estimate a model of self-assessed health as a function of objective health measures *m*, e. g. the utilisation of health care or physical and mental well-being as well as personal characteristics *x* like age and education (cf. Disney, Emmerson & Wakefield (2006)). First, we can write the unobservable health status as a function of *x* and *m* and unobservables u_{it}:

$$\eta_{it} = x_{it}^{'} \beta + m_{it}^{'} \gamma + u_{it} \qquad \text{(A.1)}$$

Instead of η_{it}, the categorical variable self-assessed health h_{it} is observed in the data set. This variable may be measured with a reporting error. Hence, the latent health stock h_{it}^{*} as the counterpart of the observed self-assessed health is a function of the unobservable health status η_{it} and a reporting error ε_{it} as well:[11]

$$h_{it}^{*} = \eta_{it} + \varepsilon_{it} \qquad \text{(A.2)}$$

The latent health variable can be linked to the dichotomous indicator h_{it} using the following observation mechanism:

$$h_{it} = j\,, \quad \text{if} \quad \mu_{j-1} < h_{it}^{*} < \mu_{j}\ , j = 1\ , \ldots , 5 \qquad \text{(A.3)}$$

Equation (A.3) shows that our observable health variable takes the value *j* if the latent health stock lies between the two thresholds μ_{j-1} and μ_{j}. Combining the observation mechanism with equation (A.1), the model can be estimated using ordered probit techniques. Using the predicted values from these estimations, we can normalise the health stock using a z-transformation. This yields to a health capital stock with a zero mean and a constant variance

[11] Disney et al. (2006) assume that the error terms in A.1 and A.2 are uncorrelated.

of one. Furthermore, positive values of our health capital stock variable indicate that the respondent's health is above the sample mean in this period.

In the estimation at hand, we use the variables physical functioning, role physical, bodily pain, vitality, social functioning, role emotional, and mental health. These are elements of the SF-12v2 indicators mentioned above (for a detailed description see Andersen et al. (2007)). In addition, interaction terms combining the log equivalent household net income with the four relevant age categories are used as exogenous variables (inc_age). This is done to control for age-dependent income effects on health assessment. The descriptive statistics are shown in table A. Table B refers to the estimation results.

<table A around here>

<table B around here>

These results presented in Table B are then used to calculate the health capital stock as a linear prediction. Together with the estimated cut-points, this prediction is taken to generate a new self assessed health variable. As figure A shows, category 1 (very good health status) seems to be over-reported in the original *SAH* variable. Moreover, the movements between the five categories are reported in table C.

<figure A around here>

<table C around here>

In a last step, we normalize the continuous prediction of *SAH* by using a z-transformation. As a result, we get a new health capital stock with a zero mean and a constant variance of one with positive values indicating a better health stock. For each respondent with a health stock above average, our binary indicator used in the multivariate probit model then takes the value one while for individuals with a low health capital stock the variable takes the value zero.

Figure 1: Health production as multi-level process

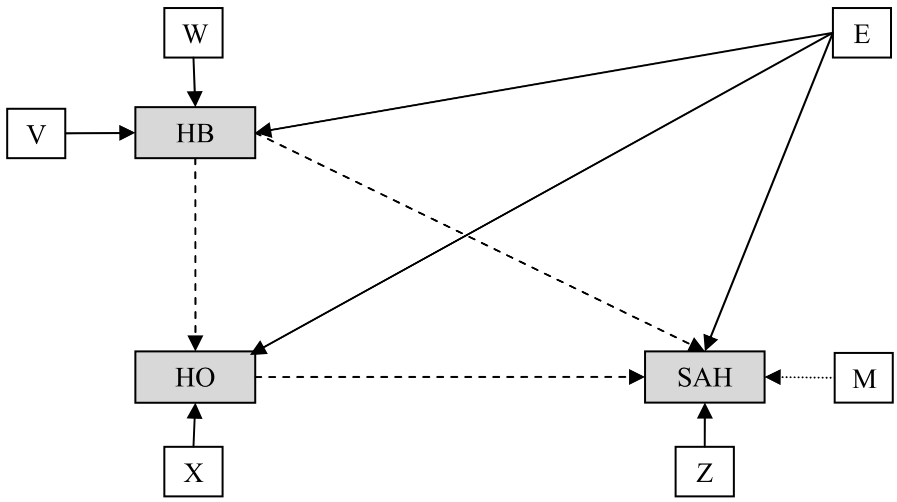

Table 1: Description of the variables

Endogenous variables	
health	self-assessed health above-average
overweight	overweight in terms of age-adjusted BMI yes/no
smoker	tobacco consumption yes/no
fast food	does not so much follow a health conscious diet yes/no
alcohol	drinks alcohol regularly yes/no
Predisposing variables	
age 17-29	respondent 17 to 29 years old yes/no (reference group)
age 30-39	respondent 30 to 39 years old yes/no
age 40-49	respondent 40 to 49 years old yes/no
age 50-64	respondent 50 to 64 years old yes/no
age $\geq$ 65	respondent older than 64 years yes/no
female	female yes/no
partnership	living together with a partner yes/no
children	at least one child younger than 16 years in household yes/no
Eastern Germany	living in Eastern Germany yes/no
foreigner	nationality not German yes/no
Socioeconomic variables	
income	log equivalent household net income per month in €
economic worries	strong worries about own economic situation yes/no
civil servant	civil servant yes/no
blue-collar	blue-collar yes/no
unemployed	long-term unemployment in 2005 and unemployed at the time the survey was conducted in 2006 yes/no
working hours 1-21	1-21 hours effectively worked per week yes/no
working hours 22-42	22-42 hours effectively worked per week yes/no
working hours >42	more than 42 hours effectively worked per week yes/no
O-level	first public examination in secondary school yes/no
high school	general qualification for university entrance yes/no
university	university degree yes/no
education	currently in some sort of education yes/no
renovation	house is at least partly in need of renovation yes/no
private health insurance	fully private insured yes/no
supplemental insurance	private supplemental health insurance yes/no

risk	prepared to take risks (0 – risk aversion to 10 – fully prepared)

Table 2: Descriptive statistics (n=9976)

Endogenous variables	Mean	Std. Dev.
alcohol	0.1595	0.3661
smoker	0.2910	0.4542
fast food	0.5023	0.5000
overweight	0.3597	0.4799
health	0.5657	0.4957
Predisposing variables		
age 30-39	0.1859	0.3891
age 40-59	0.2049	0.4036
age 50-64	0.2302	0.4210
age ≥ 65	0.2309	0.4214
female	0.5255	0.4994
partnership	0.7119	0.4529
children	0.2885	0.4531
Eastern Germany	0.2877	0.4527
foreigner	0.0835	0.2767
Socioeconomic variables		
income	7.2453	0.4666
economic worries	0.2608	0.4391
civil servant	0.1368	0.3437
blue-collar	0.1513	0.3583
unemployed	0.0348	0.1832
working hours 1-21	0.0679	0.2515
working hours 22-42	0.2732	0.4456
working hours > 42	0.1829	0.3866
O-level	0.3204	0.4666
high school	0.1125	0.3160
university	0.1728	0.3781
education	0.0753	0.2639
renovation	0.2863	0.4520
private health insurance	0.1078	0.3101
supplemental insurance	0.1221	0.3274
risk	4.7231	2.2633

Table 3: Comparison of additional exclusion restrictions for the health equation

Exclusion	I. partnerschip, children, renovation	II. partnership, children	III. renovation
AIC	54722.54	54705.98	54691.66
BIC	55630.74	55621.38	55614.28

Table 4: Estimation results: recursive multivariate probit model (N=9976)

	(1) smoker		(2) alcohol		(3) fast food		(4) overweight		(5) health	
	Coefficient/p-value		Coefficient/p-value		Coefficient/p-value		Coefficient/p-value		Coefficient/p- value	
Endogenous variables										
overweight									-0.1888	(0.636)
smoker							-0.0991	(0.482)	0.6680	(0.000)***
alcohol							0.6734	(0.001)***	-0.2782	(0.211)
fast food							0.8444	(0.000)***	-0.7384	(0.021)**
Predisposing variables										
age 30-39	-0.0651	(0.240)	-0.0144	(0.837)	-0.1673	(0.002)***	0.2046	(0.000)***	-0.5142	(0.000)***
age 40-49	-0.0385	(0.485)	0.2358	(0.001)***	-0.2329	(0.000)***	0.1318	(0.010)***	-0.9196	(0.000)***
age 50-64	-0.3172	(0.000)***	0.1288	(0.070)*	-0.3796	(0.000)***	0.1972	(0.001)***	-1.3445	(0.000)***
age ≥ 65	-0.8278	(0.000)***	0.2203	(0.007)***	-0.5900	(0.000)***	-0.0042	(0.963)	-1.7268	(0.000)***
female	-0.1524	(0.000)***	-0.7798	(0.000)***	-0.4766	(0.000)***	-0.0015	(0.982)	-0.2901	(0.000)***
partnership	-0.2210	(0.000)***	-0.0519	(0.191)	-0.0892	(0.006)***	0.1920	(0.000)***	-0.0419	(0.427)
children	-0.0770	(0.039)**	0.0160	(0.718)	-0.0345	(0.338)	0.0290	(0.417)	-0.2292	(0.000)***
Eastern Germany	-0.0785	(0.021)**	0.0107	(0.780)	0.0020	(0.949)	0.0426	(0.166)	-0.0992	(0.003)***
foreigner	0.1257	(0.014)**	-0.3926	(0.000)***	0.0401	(0.418)	0.0987	(0.047)**	-0.0227	(0.671)
Socioeconomic variables										
income	-0.2032	(0.000)***	0.1983	(0.000)***	-0.0410	(0.246)	-0.1622	(0.000)***	0.2334	(0.000)***
economic worries	0.2194	(0.000)***	-0.0698	(0.079)*	0.0628	(0.047)**	0.1138	(0.001)***	-0.3521	(0.000)***
civil servant	0.0227	(0.616)	0.0198	(0.705)	-0.0170	(0.701)				
blue-collar	0.0767	(0.081)*	0.0434	(0.388)	0.0991	(0.019)**				
unemployed	0.4085	(0.000)***	0.2042	(0.030)**	0.1484	(0.045)**				
working h. 1-21	0.1475	(0.027)**	-0.1081	(0.203)	-0.0225	(0.741)				
working h. 22-42	0.1566	(0.002)***	0.0219	(0.722)	0.1040	(0.068)*				
working h. >42	0.2173	(0.000)***	-0.0086	(0.891)	0.2290	(0.000)***				
O-level	-0.1128	(0.001)***	0.1119	(0.006)***	-0.1201	(0.000)***	-0.0594	(0.091)*	0.1231	(0.002)***
high school	-0.3204	(0.000)***	0.1674	(0.004)***	-0.3260	(0.000)***	-0.1293	(0.043)**	0.1185	(0.099)*
university	-0.4490	(0.000)***	0.2020	(0.000)***	-0.3233	(0.000)***	-0.2337	(0.000)***	0.2908	(0.000)***
education	-0.2964	(0.000)***	-0.2978	(0.000)***	0.0362	(0.599)				
renovation	0.0735	(0.029)**	0.0693	(0.070)*	0.1355	(0.000)***	-0.0056	(0.904)		

private health insurance	-0.0821	(0.109)	0.1293	(0.015)**	-0.1456	(0.001)***	-0.0312	(0.519)		
supplemental insurance	0.0177	(0.682)	0.1462	(0.003)***	-0.0923	(0.025)**	0.0442	(0.290)		
risk	0.0601	(0.000)***	0.0165	(0.044)**	-0.0253	(0.001)***				
constant	1.1073	(0.000)***	-2.3961	(0.000)***	1.0617	(0.000)***	0.0985	(0.772)	-0.0843	(0.838)
Log pseudo Likelihood	-27217.83									
Correlation coefficients	ρ_{21}= 0.11***									
	ρ_{31}= 0.26***		ρ_{32}= 0.11***							
	ρ_{41}= -0.11		ρ_{42}= -0.47***		ρ_{43}= -0.37***					
	ρ_{51}= -0.35***		ρ_{52}= 0.21		ρ_{53}= 0.39**		ρ_{54}= -0.13			
LR Test[A] $\chi^2(10)$	324.858	(0.000)***								
AIC	54691.66									
BIC	55614.28									

*** significant at the 1-%-level

** significant at the 5-%-level

* significant at the 10-%-level

[A] Likelihood ratio test of no joint significance of the covariance parameters ρ_{jk} between the equations' error terms.

Table A: Descriptive statistics (n=9976)

Additional variables	Mean	Std. Dev.	Min	Max
SAH	2.68	0.96	1	5
handicap	0.13	0.34	0	1
hospital	0.12	0.33	0	1
1-2 doctor visits	0.36	0.48	0	1
3-4 doctor visits	0.18	0.38	0	1
at least 5 doctor visits	0.15	0.36	0	1
physical functioning	49.23	10.28	27.25	58.35
role physical	49.35	10.21	21.92	59.72
bodily pain	49.36	10.22	23.00	59.85
vitality	49.43	9.97	26.82	70.60
social functioning	49.47	10.36	14.69	57.12
role emotional	49.53	10.14	13.34	58.08
mental health	49.80	10.23	19.73	68.58
inc_age 30-39	1.35	2.83	0	8.75
inc_age 40-49	1.49	2.95	0	9.21
inc_age 50-64	1.68	3.09	0	9.35
inc_age $\geq$ 65	1.66	3.04	0	9.09

Table B: Estimation Results

	SAH	
	Coefficient / p-value	
age 30-39	1.1095	0.018)**
age 40-49	1.5776	(0.000)***
age 50-64	0.5359	(0.126)
age $\geq$65	0.8060	(0.048)**
female	-0.1289	(0.000)***
partnership	0.0651	(0.024)**
children	-0.0610	(0.064)*
O-level	-0.0278	(0.334)
high school	-0.0104	(0.806)
university education	-0.0788	(0.024)**
handicap	-0.2777	(0.000)***
hospital	0.1447	(0.000)***
1-2 doctor visits	0.1972	(0.000)***
3-4 doctor visits	0.4151	(0.000)***
at least 5 doctor visits	0.6320	(0.000)***
physical functioning	-0.0365	(0.000)***
role physical	-0.0212	(0.000)***
bodily pain	-0.0294	(0.000)***
vitality	-0.0215	(0.000)***
social functioning	-0.0073	(0.000)***
role emotional	0.0009	(0.635)
mental health	-0.0134	(0.000)***
private health insurance	0.0231	(0.571)
supplemental insurance	-0.0760	(0.031)**
inc_age 30-39	-0.1276	(0.047)**
inc_age 40-49	-0.1569	(0.006)***

inc_age 50-64	0.0080	(0.866)
inc_age ≥ 65	-0.0313	(0.578)
Log pseudo-likelihood	-9334.54	
Wald test ($\chi^2(28)$)	5348.83	***
Pseudo R^2	0.3048	

*** significant at the 1-%-level

** significant at the 5-%-level

* significant at the 10-%-level

Figure A:

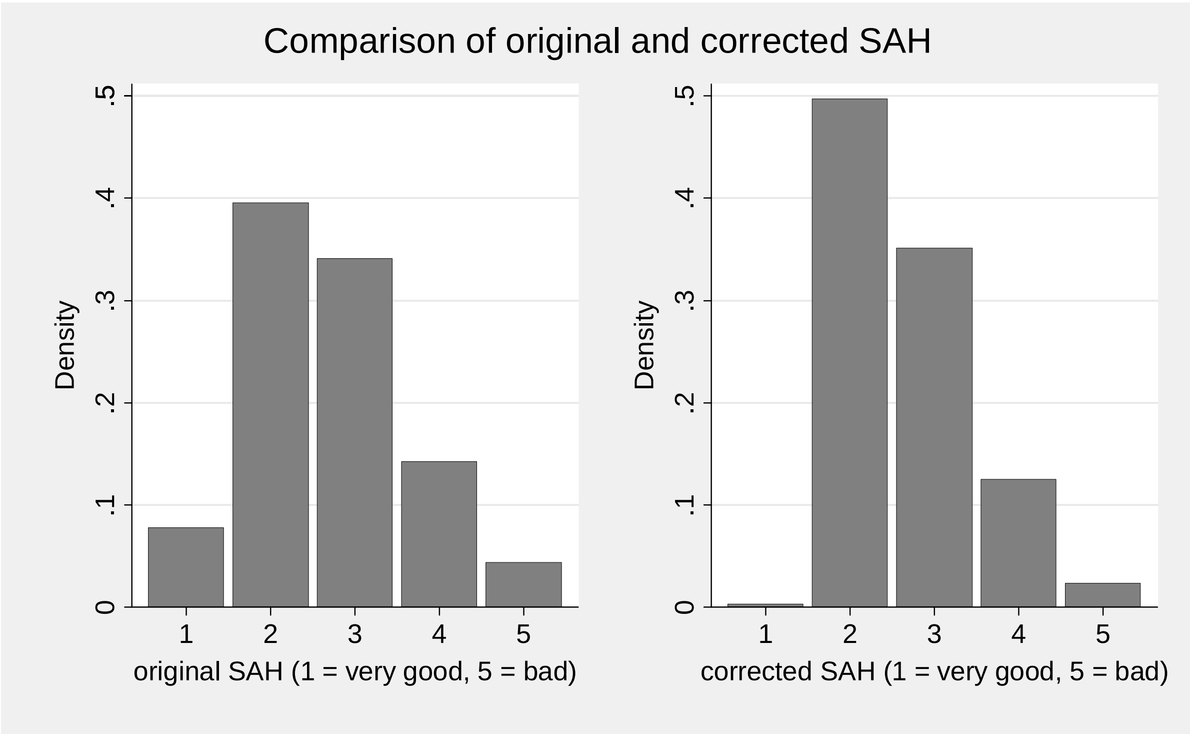

Table C: Comparison of SAH and estimated health capital

		Original SAH very good (1)	Good (2)	Satisfactory (3)	Poor (4)	Bad (5)	total
Corrected SAH	(1)	14	18	0	0	0	32
	(2)	718	3.077	1.095	63	6	4.959
	(3)	43	828	1.984	610	36	3.501
	(4)	0	25	315	679	230	1.249
	(5)	0	0	6	68	161	235
total		775	3.948	3.400	1.420	433	9.976