New estimates of the demand for health: results based on a categorical health measure and Swedish micro data

Ulf-G Gerdtham* and Magnus Johannesson,

Department of Economics, Stockholm School of Economics, Stockholm, Sweden

Working Paper Series in Economics and Finance No. 205

November 1997

Abstract: In this paper we estimate a ”Grossman” model of demand for health based on Swedish micro data. The data set consists of a random sample of over 5,000 individuals takern from the Swedish adult population. Health capital is measured by a categorical measure of overall health status, and an ordered probit model is used to econometrically estimate the demand-for-health equation. The results are consistent with the theoretical predictions and show that the demand for health increases with income and education and decreases with age, overweight, urbanisation and being single.

Key words: demand for health, health production, human capital, health status.

JEL-Classification: I12, J24, D12

*Corresponding author:
Department of Economics
Stockholm School of Economics
Box 6501
S-113 83 Stockholm
Sweden

Phone: +46 8 7369283; Fax: +46 8 302115; E-mail: heug@hhs.se

Draft 1997-10-06

1. Introduction

The determinants of health constitute an issue of vital importance to health policy, e.g. in establishing the effects of various non-medical factors on health such as unemployment or income. The first economist to present a formal model on the determinants of health was Grossman (1972a,b). Drawing on the theory of human capital by Becker (1965), Grossman constructed a model where individuals use medical care and their own time to produce health. Individuals were assumed to invest in health production until the marginal cost of health production equalled the marginal benefits of improved health status. The health status was assumed to affect utility both indirectly through raising labour income and directly by assuming that individuals value good health per se.

The seminal demand-for-health model by Grossman (1972a,b) has become a corner-stone in the field of health economics. Despite this, relatively little empirical work has been carried out on the demand-for-health model (Grossman 1972a; Cropper 1981; Wagstaff 1986,1993; Pohlmeier & Ulrich 1995). The work that has been done has yielded conflicting results, with many that have been inconsistent with the predictions of the model (Wagstaff 1993). There is thus a need for further empirical work on the demand-for-health model.

The aim of this paper is to estimate a demand-for-health equation using Swedish micro data. In previous studies the measure of health capital has been constructed from various health problems. Here we instead measure health capital by a categorical measure of overall health status (bad health, fair health or good health). This measure offers potentially important advantages compared to previous studies, since it is not necessary to determine weights for various health problems and the size of the regression coefficients can be interpreted in a meaningful and intuitive way (as the change in the probability of good health). As independent variables we use a host of variables hypothesized to affect the demand for health. In Section 2 below we present a theoretical model of the demand for health. In Section 3 the data and the variables used are described and in Section 4 the estimation methods are outlined. Section 5 reports the results and the paper ends with some concluding remarks in Section 6.

2. The theoretical model

In this section we present a theoretical model of the demand for health based on the model in Johansson (1995).[1] It is assumed that the individual behaves as if he/she maximizes utility in each of T periods, utility being a function of private goods, leisure time and health. For the sake of simplicity, the utility function is assumed to be separable in time and is written as:

$$U= \Sigma_t \rho^{t-1} U_t(Z_t,L_t,H_t) \quad (1)$$

where $\rho=(1+\theta)^{-1}<1$ is a discount factor, θ is the rate of time preference, Z_t is a vector of private goods consumed in period t with t=1,....T, L_t is leisure time in period t, and H_t is the health status in period t. The utility function is strictly increasing in each of its arguments. At time 0 the individual inherits an initial stock of health capital H_0, and the health status then evolves over time according to the relationship:

$$dH_t=I_t- \delta_t(A_t)H_t \quad (2)$$

where dH_t is the net investment in health in period t, I_t is the gross investment in health in period t, δ_t is the rate of depreciation of health in period t, and A_t is a vector of exogenous variables that affect the rate of depreciation such as age, sex, and environmental variables. The gross investment is assumed to be produced using medical care (M) and own time (L_{Ht}) as inputs, which gives rise to the following health production function:

$$I_t=I_t(M_t,L_{Ht},E_t) \quad (3)$$

[1] The model in this section differs slightly from the original Grossman model (1972a,b). In the original Grossman model the stock of health only affects utility indirectly through the number of healthy days in a period that is a function of the stock of health. In our model the stock of health (the health status) enters the utility function directly. This recognizes that there is a continuum of health states rather than only the two health states: healthy and unhealthy (as implied by letting only the number of healthy days affect utility). In our model leisure time is also an argument in the utility function, rather than being an argument in a production function to produce private goods (Z). This simplifies our model, since a production function for private goods is not needed.

where E_t is the exogenous education level that is assumed to affect the productivity of producing health. The individual is assumed to have access to a perfect capital market, allowing him or her to freely transfer income between the periods at a fixed market rate of interest that is assumed to equal the rate of time preference. This leads to the following budget constraint:

$$\Sigma_t \rho^{t-1}(Y_t+W_t(L_{tot}-L_t-L_{Ht}-L_{st}))=\Sigma_t \rho^{t-1}(p_t Zt+P_t Mt) \qquad (4)$$

where Y_t is non-wage income in period t, W_t is the exogenous wage rate in period t, L_{tot} is the total time per period, L_{st} is the sick time in period t, p_t is the price of private goods in period t, and P_t is the price of medical care in period t. Sick time in period t (L_{st}) is a function of the health status in period t and is defined as the amount of time in period t that the individual is unable to work due to illness. Maximization of the life time utility function subject to the health production function and budget constraint yields the following demand-for-health function at time t:

$$H_t=H_t(p,P,Y,W,A,E,\theta,H_0) \qquad (5)$$

where p is a vector of prices of private goods in periods 1,....T, P is a vector of medical care prices in periods 1,....T, Y is a vector of non-wage incomes in periods 1,.....T, W is a vector of wage rates in periods 1,....T, A is a vector of variables in periods 1,.....T affecting the rate of depreciation of health, and E is a vector of education levels in periods 1,......T, θ is the rate of time preference, and H_0 is the initial inherited stock of health capital at time 0. To estimate the demand function in equation (5) a functional form has to be specified. Assuming a linear functional form and that p, P, Y, W, A and E do not vary between periods, this leads to the following demand-for-health equation:

$$H_t = b_1 + b_2*p_t + b_3*P_t + b_4*Y_t + b_5*W_t + b_6*A_t + b_7*E_t + b_8*\theta + b_9*H_0 + u_{1t} \qquad (6)$$

In equation 6 u_{1t} is an error term with zero mean and constant variance and b_1-b_9 are coefficients to be estimated. We would expect the demand for health to increase with an increase in the price of private goods

($b_2>0$), i.e. as health production becomes relatively cheaper compared to private goods the individual will increase the demand for health. For the price of medical care we expect a negative sign ($b_3<0$), since the individual can be expected to purchase less medical care as the price increases. An increase in the non-wage income can be expected to increase the demand for health ($b_4>0$). For a ceteribus paribus increase in the wage rate the sign is indeterminate a priori. An increased wage will increase the price of own time in the production of health, but also increase the labour income and the value of investments in health. For the variables affecting the rate of depreciation (A_t) we expect negative signs for variables that increase the rate of depreciation and positive signs for variables that decrease the rate of depreciation. For education we expect a positive sign ($b_7>0$), since education is assumed to increase the productivity of producing health. The coefficient for the rate of time preference can be expected to be negative ($b_8<0$), since a higher discount rate decreases the present value of investments in health. The coefficient reflecting the inherited stock of health capital can be expected to be positive ($b_9>0$). Below we specify the data used to estimate equation (6).

3. Data

The empirical analysis of the demand for health is based on data from a probability sample of the Swedish population, the Level of Living Survey (LNU) from 1991. The **full sample** (employed and unemployed) consists of 6,773 individuals, between the ages 18 and 76 years. After correcting for missing values, the sample was reduced to 5,174 individuals for the full sample and 3,184 individuals in the employed sample. The survey contains data on, for instance, morbidity and different socioeconomic variables. Further details are given in Levnadsnivåundersökningen (1991) and Fritzell and Lundberg (1994).

3.1 Dependent variable (H_t)

In Table 1 the variables used are defined and in Table 2 summary statistics for the variables are given for both the full sample and the employed sample. The dependent variable is the stock of health (H_t), which is measured by a categorical health measure. In the categorical health rating question the individual's rated their own current health status on a three-point scale (0=bad health, 1=fair health, 2=good health). This type of categorical health measure has been shown to capture important information about the individual's health

(Connelly et al. 1989) and to be an important predictor of mortality (Wannamethee & Shaper 1991; Kaplan & Camacho 1983; Idler & Kasl 1991).

TABLE 1 IN HERE

TABLE 2 IN HERE

3.2 Independent variables

The independent regressor variables derived from eq (6) are outlined below.[2]

3.2.1 The non-wage income (Y) and the wage rate (W)

The Grossman model (and the theoretical model above) is based on a working population, where the wage rate measures the opportunity cost of time. In the non-working population we have no information about the opportunity cost of time. For the full sample we therefore include one income variable (gross income) rather than non-wage income and the wage-rate. We use dummy variables rather than a continuous income variable to avoid making assumptions about the functional relationship between income and health status. Gross annual income is measured by three dummy variables: INC2=1 if the income is in the second quartile (61,500-116,300), INC3=1 if the income is in the third quartile (116,300-161,700) and INC4=1 if the income is in the fourth quartile (>161,700). The reference category is individuals who are in the first (poorest) quartile of the gross income distribution (<61,500). The source of the gross annual income data is the National Income Tax Statistics, linked to the LNU data. Using gross individual income as an independent variable introduces a potential endogeneity problem. According to the theoretical model health status affects labour income (through reducing sick time), and this leads to an endogeneity problem between income and health. Since workers in Sweden have paid sick-leave the endogeneity problem will, however, be reduced.[3] We also estimate the model only for the individuals who are working (the employed sample). In this estimation we include a dummy

---

[2] It was not possible to include the price of private goods (p), the price of medical care (P) and the rate of time preference (θ; assumed to equal the market rate of interest) in the estimated demand-for-health equation, since they do not vary between individuals.

[3] There could still be an endogeneity problem if the health status affects the wage rate, which seems possible. In such a case the theoretical model would, however, be rejected since it is based on an exogenously given wage-rate (and the problem could thus

variable for non-wage income defined as 1 if the individual has some declared taxable property, and 0 otherwise. We also include the wage rate. The gross wage rate per hour is measured by three dummy variables: WAGE2=1 if the wage rate is in the second quartile (64-74), WAGE3=1 if the wage rate is in the third quartile (74-89) and WAGE4=1 if the wage rate is in the fourth quartile (>89). The reference category is individuals who are in the first quartile of the wage rate distribution (<64). All ”monetary” variables are given in Skr (Swedish Kronor; exchange rate September 1997 $1=Skr 7.80).

3.2.2 The variables affecting the rate of depreciation in health (A)

We include age as a variable that affects the rate of depreciation, since health status decreases with age. Rather than impose a functional form on the relationship between health and age we again conservatively use three 0-1 dummies for age groups (AGE2, AGE3 and AGE4). Gender is also included in the model and is represented by a 0-1 dummy for male (MALE). The rate of depreciation is assumed to be higher for men, since they have a lower life-expectancy. We also include a dummy variable for overweight measured as a body mass index over 30 (BMI>30). We expect overweight to increase the rate of depreciation. We also include one dummy variable for if the individual is not married or cohabitant (SINGLE) and one dummy variable for if the individual is currently unemployed (UNEMPL). Both these variables are expected to increase the rate of depreciation. We also include two dummy variables for urbanisation (SMALLCITY, BIGCITY). We expect urbanisation to increase the rate of depreciation.

3.2.3 Education (E)

We include two dummies for the education of the individual (EDUC2, EDUC3).[4] We expect the productivity of producing health and thus the demand for health to increase with higher education.

3.2.4 The initial inherited stock of health capital (H_0)

We use a proxy variable for the initial inherited stock of health capital. We use a dummy variable for if the parents or siblings of the respondent have had any health problems (HPROBFMS). Since we expect the

---

not be solved by including non-wage income and the wage rate rather than total income as regressors). Due to the possible endogeneity problem of income we also estimated the demand for health model without the income variable, see footnote 5.

demand for health to increase with the initial inherited stock of health capital we expect a negative sign for this variable.

4. Estimation methods

In estimating the demand-for-health model we have to take into account the fact that our health measure is an ordered response with three categories (0=bad health, 1=fair health, 2=good health). An appropriate tool for analyzing such ordered categorical data is the ordered probit model (for references see Amemiya (1981); Cameron & Trivedi (1986); Greene (1993)). Let h_i^* be a continuous, latent variable which could be interpreted as representing the health of an individual on a continuous scale. We assume a linear dependence between the latent variable h_i^* and X_i, β and ε_i:

$$h_i^* = \beta' X_i + \varepsilon_i, \; \varepsilon_i \sim N(0,\sigma^2),$$

The variable h_i^* defines a variable h_i which is related to the above mentioned categories in the following way:

$$h_i = \begin{cases} 0 \text{ if } h_i^* \leq \theta_0 \\ 1 \text{ if } \theta_0 < h_i^* \leq \theta_1 \\ 2 \text{ if } \theta_1 < h_i^* \end{cases}$$

where θ_i = 0,1, are unobservable thresholds. Denoting the cumulative density function of the standard normal distribution as above (Φ), it follows that the probabilities of an individual for each category are given by:

$$\text{Prob}[h_i=0] = \Phi[\mu_0 - \alpha' x],$$

$$\text{Prob}[h_i=1] = \Phi[\mu_1 - \alpha' x] - \Phi[\mu_0 - \alpha' x],$$

[4] The advantage of education categories over a continuous education variable reflecting years of schooling is that education categories do not suffer from the problem that an additional year of schooling at, for example, the high-school level might have a

$$\text{Prob}[h_i=2] = 1- \Phi[\mu_1 - \alpha' x],$$

with $\beta = \alpha/\sigma$ and $\mu_{j-1}/\sigma = \theta_j$, 0,1, i.e. note that only the ratios α/σ and μ_{j-1}/σ are estimable (Dustman 1996). If the regression contains a constant term, the full set of coefficients is not identified. A common normalization is to set μ_0= 0 which means that the estimated coefficients μ_i, i=1 represent the differences in the respective thresholds: $\mu_i = \mu_i - \mu_{i-1}$ (Greene 1995; Dustman 1996). Greene (1993) points out that the interpretation of the estimates is not straightforward. A positive estimate indicates that an increase in the respective variable shifts weight from category 0 into category 2, which means that the probability of category 2 increases and the probability of category 0 decreases.

5. Estimation results

Our statistical results are reported in two parts: the full sample estimations (working and non-working population) and the employed sample estimations. In Table 3 we report the ordered probit estimated effects of the covariates on the categorical health measure for both the full sample and the employed sample. The goodness of fit values (pseudo R^2) are about 0.10 for the full sample and 0.07 for the employed sample. In order to be able to interpret the size of the regression coefficients, we show in Table 4 the predicted probability of good health for each category of the dummy variables, at the mean level of all other explanatory variables for both samples.

TABLE 3 IN HERE

TABLE 4 IN HERE

5.1 Full sample estimations

The estimated effects of income on the demand for health are clearly positive, i.e. the estimated effects of INC3-4 are significant (P<0.01). These effects tend to increase with increasing income and the estimated effect

---

very different effect compared with an additional year at the university level.

of INC4 is significantly higher than the effect of INC3 (P<0.01). The estimated probability of being in good health increases from 0.75 in the lowest income category to 0.87 in the highest income category.[5]

As expected age significantly decreases the demand for health, i.e. the estimated effects of AGE2-4 are negative and significant (P<0.01). The estimated effect of AGE4 is significantly lower than the corresponding effect of AGE2 (P<0.01), but not significantly lower than AGE3. The predicted probability of good health is 0.93 in the youngest age-group and 0.63 in the oldest age-group. The estimated effect of gender (MALE) is negative and significant (P<0.10).

The estimated effect of being single (SINGLE) is as expected negatively significant (P<0.01). Being single decreases the predicted probability of good health from 0.82 to 0.77. The estimated effect of unemployment is negative as expected, but not significant. The estimated effect of the variable for health problems in the family (HPROBFMS) is as expected negatively significant (P<0.10). The estimated effect of overweight (BMI>30) is negatively significant (P<0.01) and the predicted probability of being in good health decreases from 0.81 to 0.69 with overweight. The estimated effect of living in a big city (Stockholm, Gothenburg or Malmö) is significantly negative (P<0.01) but the estimated effect of living in cities larger than 30,000 inhabitants is not significant.

Finally, education has a positive effect on health, indicating that individuals with a higher education are more efficient producers of health. The two education variables (EDUC2 and EDUC3) are positive and significant (P<0.01). The predicted probability of being in good health is 0.89 for individuals with university education and 0.78 for individuals with less than high school education.

5.2 Employed sample estimations

[5] It might be expected that income (INC2-4) will be endogeneous to health. If this is the case, then the estimated coefficients on income will be biased, along with the coefficients of all other correlated regressors. Unfortunately, the lack of instruments precludes some formal tests of possible endogeneity. However, estimation of the reduced form equation, i.e. without the dummies for gross income, showed that at least our qualitative results of the original models presented in Table 3 were robust, i.e. the significance and signs of the estimated effects were only affected in one case: the dummy for MALE is positive and not significant in the reduced form equation. These estimations are available from the authors upon request.

The estimated effect of the wage rate on the demand for health is positive and significant, i.e. WAGE2-4 are all positive and significant (P<0.05) and the estimated effect tends to increases as the wage rate increases. The predicted probability of being in good health is 0.81 in the lowest wage rate category and 0.90 in the highest wage rate category. The estimated effect of non-wage income on the demand for health is also positive and significant (P<0.01).[6]

Age significantly decreases health also among employed people, i.e. the estimated effect of AGE3 is significant (P<0.01) but the effect of AGE2 is not significant (note that AGE4 is not included). As in the full sample, the estimated effects of gender (MALE) and being single (SINGLE) are negatively significant (p<0.05).

The estimated effects of the dummy for health problems in the family (HPROBFMS) and living in big cities (BIGCITY) are negative, as was the case for the full sample, but they are not significant. The estimated effect of overweight (BMI>30) is negatively significant also for the employed sub-sample and being overweight decreases the predicted probability of good health from 0.87 to 0.76. Finally, education has a positive effect on health also in the employed sample. The two education dummy variables (EDUC2 and EDUC3) are positive and significant (P<0.01). The predicted probability of being in good health is 0.91 for individuals with university education and 0.84 for individuals with less than high school education.

6. Concluding remarks

Overall our results confirm the predictions of the demand-for-health model. The demand for health decreases with age and is lower for males than for females, and the demand for health increases with income and education. In the employed sample the demand for health decreases with age and is lower for males than females, and the demand for health increases with the wage rate, non-wage income and education. Overweight and being single also had a significant negative effect on the demand for health in both samples. We furthermore found that living in big cities (Stockholm, Gothenburg or Malmö) had a negative effect on health status, although this effect was not significant in the employed sample. Our proxy variable for the initial inherited stock of health capital had the expected negative sign in both samples, but was not significant in the

[6] Since it is possible that also the wage rate is endogenous to health, see footnote 3 and 5, we also re-estimated the reduced form

employed sample. However, we failed to find a significant impact of unemployment, although it had the expected negative sign.

The most recent empirical work that we are aware of which estimates demand-for-health equations is that of Wagstaff (1993). He estimated demand-for-health equations with income, education, sex and age as explanatory variables. He estimated two equations in two separate age-groups (under and over 41 years). Age had the wrong sign (not significant) in the under 41s' equation, but had the expected negative sign and was significant in the over 41s' equation. Income and education had the expected signs in both equations, but income was not significant in the under 41s' equation and education was not significant in the over 41s' equation. Our results for age, income and education are largely consistent with those of Wagstaff (1993), but provide stronger support for the demand-for-health model. Age, education and income are all significant with the expected signs in our estimations. We are also able to estimate the effect of these variables in an easily interpretable way, as the change in the predicted probability of good health.

Finally, it is important to note some limitations of the study. One limitation concerns the fact that there may be some omitted variables which affect the demand for health (e.g. the distance to health care services, affecting the effective price of medical care). This could lead to omitted variables bias. Another limitation concerns causality. For some variables there could be problems with reversed causality, e.g. that the health status affects income and education rather than the other way round. If this is the case, the estimated effects of income and education will be biased, along with the effects of all other correlated regressors. Unfortunately, the lack of instruments precludes any formal tests of endogeneity. In spite of these limitations we think that the results are interesting and provide some support for the demand-for-health model.

---

equation without the dummies for wage rates, and we found that all qualitiative results of the other variables hold, except for the insignificance of the male dummy, as was also the case in the reduced form equation for the full sample, see footnote 5.

References

Amemiya, T., 1981, Qualitative response model: a survey, Journal of Economic Literature 19, 481-536.

Becker, G.S., 1965, A theory of the allocation of time, Economic Journal 75, 493-517.

Bleichrodt, H. and Johannesson, M., 1997, An experimental test of a theoretical foundation for rating scale valuations, Medical Decision Making 17, 208-216.

Blumenschein, K. and Johannesson, M., 1997, The relationship between quality of life instruments, health state utilities and willingness to pay in patients with asthma, Annals of Allergy, Asthma, and Immunology, in press.

Bosch, J.L. and Hunink, M.G.M., 1996, The relationship between descriptive and valuational quality-of-life measures in patients with intermittent claudication, Medical Decision Making 16, 217-225.

Cameron, C. and Trivedi, P., 1986, Econometric models based on count data: comparisons and applications of some estimators and tests, Journal of Econometrics 1, 29-53.

Connelly, J.E., Philbrick, J.T., Smith, R., Kaiser, D.L. and Wymer, A., 1989, Health perceptions of primary care patients and the influence on health care utilization. Supplement to Medical Care 27, 99-109.

Cropper, M.L., 1981, Measuring the benefits from reduced morbidity, American Economic Review 71, 235-240.

Dustman, C., 1996, The social assimilation of immigrants, Journal of Population Economics 9, 37-54.

Gold, M.R., Siegel, J.R., Russel, L.B. and Weinstein, M.C., eds., 1996, Cost-effectiveness in health and medicine, Oxford University Press, New York.

Greene, W., 1993, Econometric Analysis, Second Ed., Macmillan Publishing Company, New York.

Greene, W., 1995, LIMDEP™ Version 7.0 User’s manual, Econometric Software, Inc.

Grossman, M., 1972a, The demand for health: a theoretical and empirical investigation, NBER, New York.

Grossman, M., 1972b, On the concept of health capital and the demand for health, Journal of Political Economy 80, 223-255.

Hornberger, J.C., Redelmeier, D.A. and Peterson, J., 1992, Variability among methods to assess patients’ well-being and consequent effect on a cost-effectiveness analysis, Journal of Clinical Epidemiology 45, 505-512.

Idler, E.L. and Kasl, S., 1991, Health perceptions and survival: do global evaluations of health status really predict mortality?, Journal of Gerontology 46, 555-565.

Johansson, P-O., 1995, Evaluating health risks: an economic approach, Cambridge University Press, Cambridge.

Kaplan, G.A. and Camacho, T., 1983, Perceived health and mortality: a nine-year follow-up of the human population laboratory cohort, American Journal of Epidemiology 117, 292-304.

Magee, L., 1990, R^2 measures based on Wald and likelihood ratio joint significance tests, The American Statistician 44, 250-253.

von Neumann, J. and Morgenstern, O., 1947, Theory of games and economic behavior, Princeton University Press, Princeton, NJ.

Pliskin, J.S., Shepard, D.S. and Weinstein, M.C., 1980, Utility functions for life years and health status, Operations Research 28, 206-224.

Pohlmeier, W. and Ulrich, V., 1995, An econometric model of the two-part decisionmaking process in the demand for health care, Journal of Human Resources 30, 339-361.

Read, J.L., Quinn, R.J., Berwick, D.M., Fineberg, H.V. and Weinstein, M.C., 1984, Preference for health outcomes: comparison of assessment methods, Medical Decision Making 4, 315-329.

Stewart, A.L., Hays, R.D. and Ware, J.R., 1988, The MOS short-form general health survey: reliability and validity in a patient population, Medical Care 26, 724-735.

Torrance, G.W., 1986, Measurement of health state utilities for economic appraisal: a review, Journal of Health Economics 5, 1-30.

Veall, M.R. and Zimmermann, K.F., 1996, Pseudo-R2 measures for some common limited dependent variable models, Journal of Economic Surveys 10, 241-259.

Wagstaff, A., 1986, The demand for health: some new empirical evidence, Journal of Health Economics 5, 195-233.

Wagstaff, A., 1993, The demand for health: an empirical reformulation of the Grossman model, Health Economics 2, 189-198.

Wannamethee, G. and Shaper, A.G., 1991, Self-assessment of health status and mortality in middle-aged British men, International Journal of Epidemiology 20, 239-245.

Table 1: Definition of the variables used in the analysis

Variable	Definition
Dependent variables	
CATEGORICAL HEALTH	Assessment of own health on a three-point scale (0=bad health, 1=fair health, 2= good health)
Independent variables	
MALE	= 1 if male
AGE1	= 1 if age is 18-34 years
AGE2	= 1 if age is 35-44 years
AGE3	= 1 if age is 45-64 years
AGE4	= 1 if age is >64 years
BMI>30	= 1 if body mass index (BMI) is higher than 30[#]
UNEMPL	= 1 if unemployed
SINGLE	= 1 if the individual is not married or cohabiting
HPROBFMS	= 1 if the parents or siblings had any health problems
EDUC1	= 1 if less than high school education
EDUC2	= 1 if high school education
EDUC3	= 1 if university education
INC1	= 1 if the gross annual income is in the first quartile of the income distribution, i.e. <61,500 Skr
INC2	= 1 if the gross annual income is in the second quartile of the income distribution, i.e. 61,500-116,300 Skr
INC3	= 1 if the gross annual income is in the third quartile of the income distribution, i.e. 116,300-161,700 Skr
INC4	= 1 if the gross annual income is in the fourth quartile of the income distribution, i.e. >161,700 Skr
WAGE1	= 1 if the gross hourly wage rate is in the first quartile of the wage distribution among employed persons, i.e. <64 Skr
WAGE2	= 1 if the gross hourly wage rate is in the second quartile of the wage distribution among employed persons, i.e. 64-74 Skr
WAGE3	= 1 if the gross hourly wage rate is in the third quartile of the wage distribution among employed persons, i.e. 74-89 Skr
WAGE4	= 1 if the gross hourly wage rate is in the fourth quartile of the wage distribution among employed persons, i.e. >89 Skr
NON-WAGE	= 1 if positive taxable and declared property
RURAL	= 1 if the individual lives on the countryside or in cities with <30,000 inhabitants
SMALLCITY	= 1 if the individual lives in cities larger than 30,000 inhabitants
BIGCITY	= 1 if the individual lives in Stockholm, Gothenburg or Malmö

[#] BMI is defined as: kg/m^2, where kg is the weight in kilograms and m is the height in meters.

TABLE 2: SAMPLE DESCRIPTIVE STATISTICS
(N=NUMBER OF OBSERVATIONS).

	TOTAL SAMPLE					EMPLOYED SUB-SAMPLE				
VARIABLES	**MEAN**	**SD**	**MIN**	**MAX**	**N**	**MEAN**	**SD**	**MIN**	**MAX**	**N**
CATEGORICAL	1.7327	0.5311	0.0000	2.0000	5174	1.8216	0.4387	0.0000	2.0000	3184
MALE	0.5027	0.5000	0.0000	1.0000	5174	0.4937	0.5000	0.0000	1.0000	3184
AGE2	0.3962	0.4892	0.0000	1.0000	5174	0.5101	0.5000	0.0000	1.0000	3184
AGE3	0.3199	0.4665	0.0000	1.0000	5174	0.3775	0.4848	0.0000	1.0000	3184
AGE4	0.1446	0.3517	0.0000	1.0000	5174	-	-	-	-	-
BMI30	0.0580	0.2337	0.0000	1.0000	5174	0.0452	0.2078	0.0000	1.0000	3184
SINGLE	0.3288	0.4698	0.0000	1.0000	5174	0.2773	0.4477	0.0000	1.0000	3184
UNEMPL	0.0446	0.2065	0.0000	1.0000	5174	-	-	-	-	-
HPROBFMS	0.1504	0.3575	0.0000	1.0000	5174	0.1448	0.3519	0.0000	1.0000	3184
EDUC3	0.5176	0.4997	0.0000	1.0000	5174	0.5606	0.4964	0.0000	1.0000	3184
EDUC2	0.1637	0.3700	0.0000	1.0000	5174	0.1894	0.3919	0.0000	1.0000	3184
EDUC1	0.0750	0.2634	0.0000	1.0000	5174	0.0948	0.2931	0.0000	1.0000	3184
INC2	0.2486	0.4322	0.0000	1.0000	5174	-	-	-	-	-
INC3	0.2820	0.4500	0.0000	1.0000	5174	-	-	-	-	-
INC4	0.2839	0.4509	0.0000	1.0000	5174	-	-	-	-	-
HOURW2	-	-	-	-	-	0.2726	0.4454	0.0000	1.0000	3184
HOURW3	-	-	-	-	-	0.2610	0.4392	0.0000	1.0000	3184
HOURW4	-	-	-	-	-	0.2541	0.4354	0.0000	1.0000	3184
NON-WAGE	-	-	-	-	-	0.5710	0.4950	0.0000	1.0000	3184
SMALLCITY	0.2248	0.4175	0.0000	1.0000	5174	0.2283	0.4198	0.0000	1.0000	3184
BIGCITY	0.2779	0.4480	0.0000	1.0000	5174	0.2902	0.4539	0.0000	1.0000	3184

TABLE 3: ORDERED PROBIT MAXIMUM LIKELIHOOD ESTIMATION RESULTS: DEPENDENT VARIABLE : HEALTH (COVARIATES INCLUDED).[a]

	TOTAL SAMPLE		EMPLOYED SAMPLE	
VARIABLE	COEFF.	t-VALUES	COEFF.	t-VALUES
ONE	2.458***	27.747	2.226***	17.898
MALE	-0.721E-01*	-1.672	-0.136**	-2.287
AGE2	-0.332***	-4.272	-0.121	-1.166
AGE3	-1.002***	-12.677	-0.726***	-6.861
AGE4	-1.144***	-14.092	-	
BMI>30	-0.395***	-5.515	-0.417***	-3.917
UNEMPL	-0.115E-01	-0.115	-	
SINGLE	-0.179***	-3.976	-0.139**	-2.139
HPROBFMS	-0.101*	-1.885	-0.687E-01	-0.900
EDUC2	0.358***	5.421	0.342***	3.954
EDUC3	0.484***	5.519	0.319***	2.946
INC2	0.643E-02	0.105	-	
INC3	0.171***	2.695	-	
INC4	0.444***	6.153	-	
WAGE2	-		0.148**	2.057
WAGE3	-		0.289***	3.849
WAGE4	-		0.388***	4.657
NON-WAGE	-		0.164***	2.837
REGION4	-0.365E-01	-0.713	0.468E-01	0.643
BIGCITY	-0.130***	-2.749	-0.762E-01	-1.185
μ_1	1.077***	33.205	1.057***	21.142
N	5174		3184	
Iterations completed	21		18	
-Log-L	2987.463		1478.838	
Pseudo R^{2c}	0.100		0.066	
Pseudo R^{2d}	0.121		0.063	
% corrected	0.777		0.844	

[a]*** p<.01,** p<.05,* p<.10.
[b]Upper limit equal to 1.
cPseudo R^2 = 1 - LU/LR.
[d]Pseudo R^2 = 1- exp((-2*(LU-LR/n)), where LU is the unrestricted log likelihood values and LR is the restricted log likelihood values. n is the sample size (Magee 1990).

TABLE 4: PREDICTED PROBABILITIES OF BEING IN GOOD HEALTH		
VARIABLE	**TOTAL SAMPLE**	**EMPLOYED SAMPLE**
MALE	0.79396	0.84663
FEMALE	0.81389	0.87656
AGE1	0.92971	0.92318
AGE2	0.87313	0.90418
AGE3	0.68150	0.75819
AGE4	0.62914	-
BMI>30	0.68578	0.75564
BMI=<30	0.81030	0.86645
UNEMPL	0.80096	-
NOT UNEMPL	0.80417	-
SINGLE	0.76918	0.83899
NOT SINGLE	0.81986	0.87065
HPROBFMS	0.77954	0.84900
HPROBFMS=0	0.80818	0.86451
EDUC1	0.77671	0.84035
EDUC2	0.86847	0.90952
EDUC3	0.89354	0.90565
INC1	0.75179	-
INC2	0.75383	-
INC3	0.80271	-
INC4	0.86953	-
WAGE1	-	0.80961
WAGE2	-	0.84722
WAGE3	-	0.87817
WAGE4	-	0.89689
NON-WAGE	-	0.87726
NON-WAGE=0	-	0.84062
RURAL	0.81606	0.86484
SMALLCITY	0.80619	0.87474
BIGCITY	0.77958	0.84758

ADDITIONAL TABLE NOT FOR PUBLICATION

TABLE A: ORDERED PROBIT MAXIMUM LIKELIHOOD ESTIMATION RESULTS: DEPENDENT VARIABLE : HEALTH (COVARIATES INCLUDED).[a]

	TOTAL SAMPLE		EMPLOYED SAMPLE	
VARIABLE	COEFF.	t-VALUES	COEFF.	t-VALUES
ONE	2.419***	28.570484	2.255***	18.305
MALE	0.537E-01	1.357	-0.359E-01	-0.648
AGE2	-0.217***	-2.850	-0.370E-01	-0.354
AGE3	-0.833***	-10.943	-0.613***	-5.771
AGE4	-1.086***	-13.346	-	
BMI>30	-0.418***	-5.885	-0.439***	-4.158
UNEMPL	-0.940E-01	-0.948	-	
SINGLE	-0.188***	-4.199	-0.159**	-2.454
HPROBFMS	-0.918E-01*	-1.730	-0.564E-01	-0.745
EDUC2	0.412***	6.286	0.174***	3.039
EDUC3	0.602***	6.979	0.392***	4.654
NON-WAGE	-		0.427***	4.074
REGION4	-0.189E-01	-0.372	0.467E-01	0.646
BIGCITY	-0.102**	-2.174	-0.380E-01	-0.595
μ_1	1.063***	33.196	1.047***	21.112
N	5174		3184	
Iterations completed	18		17	
-Log-L	3016.365		-1491.010	
Pseudo R^{2c}	0.092		0.058	
Pseudo R^{2d}	0.111		0.056	
% corrected	0.776		0.844	

[a]*** p<.01,** p<.05,* p<.10.
[b]Upper limit equal to 1.
cPseudo R^2 = 1 - LU/LR.
[d]Pseudo R^2 = 1- exp((-2*(LU-LR/n)), where LU is the unrestricted log likelihood values and LR is the restricted log likelihood values. n is the sample size (Magee 1990).