ECONOMETRICS AND HEALTH ECONOMICS

AGE RELATED HEALTH DYNAMICS AND CHANGES IN LABOUR MARKET STATUS

MARCEL KERKHOFS* AND MAARTEN LINDEBOOM
CERRA, University of Leiden, The Netherlands

SUMMARY

We focus on aspects of health changes, the importance of cohort effects, age related health changes and the effect of labour market status and work history on health. We moreover assess the relative importance of gradual changes and sudden shocks in health changes and the role of work status on the likelihood of experiencing a health shock. A fixed effect panel data model is estimated on two waves of a survey of Dutch elderly. We find strong differences in health outcomes for different age cohorts and gender. We also find that health deteriorates with employment and labour market history. © 1997 by John Wiley & Sons, Ltd.

Health Econ. **6**: 407–423 (1997)

No. of Figures: 7. No. of Tables: 8. No. of References: 8.

KEY WORDS — health dynamics; work history; panel data; fixed effects model

INTRODUCTION

Life expectancy is known to increase steadily over time. Dutch males born in 1900 had a life expectancy of 55 whereas younger cohorts born in the 1980s are expected to live 74 years. The Dutch Central Bureau of Statistics (1989) predicts that elderly males of 50 years and older will increase from 25% of the male population to 34% between 1990 and 2020. Females in the same age bracket will go from 30 to 38% of female population in the same 20 year period. This trend of 'greying' compounded by the long-term reduction in labour force participation of the elderly is expected to lead to major social problems in the near future.

With respect to labour force participation of the elderly, substantial attention has been given to the interrelation between health and retirement behaviour. It is generally believed that health has a dominant effect on retirement behaviour and that individual decisions regarding labour market status and health are jointly determined. This effectively means that labour market status is expected to have a direct and/or indirect effect on health and that health in turn also determines labour market status. This paper focuses on the effect of (changes in) labour market status on health.

Individual health is a dynamic concept. Health may gradually change with age and (unforeseen) events may have a sudden impact on one's health. Labour market status may play an important role in this as it is conceivable that the rate at which health changes with age may differ across labour market states. For instance, aspects of work may

*Correspondence to: M. Kerkhofs, CERRA, Department of Economics, Leiden University, P.O. Box 9521, 2300 RA Leiden, The Netherlands.
Contract grant sponsor: NESTOR (Netherlands Institute on Ageing)
Contract grant sponsor: Leiden University

CCC 1057–9230/97/040407–17 $17.50

Received 18 October 1996
Accepted 14 March 1997

affect one's health and therefore cause health to deteriorate faster over time for individuals at work. Or it may even be the case that inactivity causes bad health. Individuals may experience health shocks due to the onset of (chronic) diseases or people may simply be involved in an accident. The likelihood of the occurrence of such shocks may depend on age, work status and notably on work environment and working conditions.

The present contribution focuses on aspects of health changes, the role of age related health changes and the effect of labour market status on these. We moreover assess the relative importance of gradual changes in and sudden shocks to health and the role of work status on the likelihood of experiencing a health shock.

The issue is of great policy relevance. Health and productivity are strongly related and policies to fight early withdrawal from the labour force aim at postponing retirement by increasing the mandatory retirement age and restricting opportunities or increasing costs of early outflow for workers and firms. Moreover, health care expenditures directly depend upon the health condition of the population. As a consequence, with the greying of the population, substantial increases in the demand for health care are expected. It is therefore of central importance to assess the effect of work on health outcomes and to know how health changes with age and whether these health profiles differ across age cohorts.

The next section discusses aspects of our data, some definitions of the health measures that we use and a preliminary analysis. It is argued that results from simple cross-sectional analyses of health are misleading owing to the endogenous nature of included regressors such as labour market status and work history. A model that accounts for the endogeneity of labour market status and labour market history is presented in the subsequent section. The empirical implementation of the model and estimation results are then presented, followed by the conclusion.

DATA

Data were obtained from the first two waves of the CERRA panel survey. The CERRA panel survey is a Dutch survey that is designed specifically for the analysis of health and retirement issues and resembles the well known Health and Retirement Survey (HRS) of the Michigan Survey Centre. The first wave was held in the autumn of 1993 and consisted of 4727 households in which the head of the household was between 43 and 63 years of age. In each household both the head and partner, if available, were interviewed. Extensive information was obtained on labour market status, sources of income, labour market history, housing, health and a variety of socio-economic variables.

In the autumn of 1995, the same respondents were contacted for a second interview. Approximately 74% of the first wave respondents participated in the second wave, which resulted in approximately 3500 households. The second wave primarily focused on the changes in labour market status, income, health status and other socio-economic variables that might have occurred in the 2 year interval.

In our analyses we focus on changes in health status between October 1993 (wave I) and October 1995 (wave II). The data contain self-assessed health measures referring to general health status and self-assessed health associated with work. Furthermore, the data contain more objective measures of health status such as the number of doctor visits, responses to the well known IADL question and individual responses to the Hopkins Symptom Checklist (HSCL). In addition to this information, the survey has information on changes in health status between the two survey dates. The respondents were asked whether sudden changes in their health condition had occurred during the 2 year time span. Moreover, the causes of sudden health changes and information on the timing of the health shocks were recorded. In our analyses we use the HSCL measures from both waves of our survey and the information on the occurrence of a health shock between the two interview dates.

There is a substantive literature on biases in subjective health measures (see, for instance, Bazzoli,[1] Anderson and Burkhauser,[2] Stern,[3] Bound[4] and Kerkhofs and Lindeboom.[5] It is argued that responses concerning health may depend upon labour market status, due to economic incentives or that responses are adapted to conform to social norms. These biases may consequently seriously distort parameter estimates in behavioural models for the relationship between labour market status and health. Therefore, these well known subjective measures will not be used in our analyses.

The HSCL is a validated objective test of general health used in the medical sciences to assess the psychoneurotic and somatic pathology of patients. The test, consisting of 57 items, is known to have an excellent rate of internal consistency, meaning that the test results are highly correlated with objective medical reports on the patients' health status. The answers to these 57 items result in a mental score, a physical score and a total health score. The advantage of this HSCL measure over subjectively, self-assessed health measures is that it is free of reporting errors that may depend upon state.

Note that the information on the occurrence of a health shock between the two interview dates can also be subject to the same criticism of the subjective health measures above. The issue of (state dependent) reporting errors is that people may *exaggerate* the extent of health problems. We feel that the relevance of state dependent reporting errors is less serious in this question as it concentrates on whether or not sudden changes in health conditions have occurred, regardless of the size of the health shocks. The degree of health changes is obtained from the HSCL measures in 1993 and 1995.

The sample includes only heads of households. In Tables A1–A3 in Appendix A we present some descriptives. Table A1 reports means of the main variables that will be used in the analyses. Table A2 provides direct information on the total (untransformed) HSCL scores of 1993 by labour market status and age group. Low HSCL scores are associated with good health. Early Retirees seem to be healthier than others, whereas Disabled are by far the least healthy. Health varies over age, but the pattern appears to be at variance with *a priori* expectations. For employed workers (wage earners), for instance, health is worst for age category 2 (age ε [51,55]) and improves thereafter. There is a similar pattern for the Disabled and the Unemployed, which makes it hard to relate this pattern to differences in work status. The last column of Table A2 reports the averages of changes in health status between the two waves of our survey. Health deteriorates over time for each of the labour market states. This hints that the age pattern implied by the first four columns of Table A2 represents age cohort effects rather than pure age effects.

Table A3 presents a multivariate analysis of health status based on the 1993 wave. The table presents excerpts from OLS regression results of the (transformed and untransformed) HSCL scores on a range of covariates, of which only age and variables related to labour market status and work experience are reported. The age pattern implied is that health deteriorates up to approximately 52 and improves thereafter. Large and significant effects of labour market status are found in Table A3. Employed wage earners, Self-Employed and Early Retirees are healthier, whereas the Disabled are unhealthier. Lagged labour market status and the number of months worked in the past 2, 5 and 10 years are included to capture effects of labour market history on health outcomes. The overall picture that emerges from these simple regression results is that working improves health. Alternatively, the results could also reflect that those in good health are able to perform work and that health problems lead to withdrawal from the labour force.

The large literature on retirement behaviour seems to suggest that the causality may also run from health to work status or that health and work status are related through unobservables. In fact, the main conclusion from the large retirement literature is that health has an important effect on retirement behaviour and that in assessing the effect of health on retirement behaviour, the endogenous nature of health has to be taken into account. This would imply that the assumption of exogeneity of the included regressors is violated in the health level equation and suggests that the results reported in Table A3 should not be taken too seriously. To assess the effect of work status on health, alternative estimation procedures are required that take the endogenous interrelation between these variables into account. This will be discussed in the next section.

THE MODEL

Our main interest is in the effect of age and labour market status on changes in health. Labour market outcomes and health status may be jointly determined. There may be direct (causal) effects that run from labour market status to health. For example, stress associated with work may cause health to deteriorate faster over time or inactivity in itself may cause bad health. On the other hand, health and labour market status may be endogenously related through unobserved factors. In the context of a life cycle model, for instance, decisions regarding health and labour market are jointly taken. Those with intrinsic low rates of time preference may be more inclined to invest in future labour market positions and health than those with higher time discount rates, causing labour market status and health to be related at even more advanced ages (see, e.g., Fuchs[6]). An individual's health condition may change gradually over time or sudden events may happen that cause health to deteriorate or improve. Individuals may experience health shocks due to the onset of (chronic) diseases or people may be simply involved in an accident. The likelihood of the occurrence of such shocks may depend on age, work status and notably on work environment and working conditions.

In line with this, we can construct an equation that (omitting an index for individual variation) relates health (h_t) to a function of labour market status $f_1(L_t)$, a function of health shocks that may have occurred in the period preceding t [i.e. (t–**1**,t)], $f_2(S_t)$ and a vector of exogenous variables (X_t):

$$h_t = \alpha_0 + \alpha_1 f_1(L_t) + \alpha_2 f_2(S_t) + \beta' X_t + \gamma + \varepsilon_t \quad (1)$$

ε_t is an iid error term that is independent of $f_1(L_t)$, $f_2(S_t)$ and X_t and γ. γ is an unobserved individual time constant component associated with health status. The unobserved individual component γ exhibits elements of the initial stock of health and decisions made in the course of the life cycle regarding labour market status and health. As a consequence labour market status or rather $f_1(L_t)$ and possibly $f_2(S_t)$ may be correlated with γ. The arguments of the functions $f_1(.)$ and $f_2(.)$ are indexed by t, but it should be noted that also lagged values of L_t and S_t could be included. We return to this issue in the discussion of the empirical implementation of the model. The effect of age, birth cohort, marital status, education, etc., on the level of health will be captured by X_t.

Estimation of equation (1) using the standard random effects approach requires the model to be augmented with an explicit model for $f_1(L_t)$ [and presumably also $f_2(S_t)$]. Specification of $f_1(L_t)$ has gained substantive interest over the past two decades in the literature on retirement behaviour (for a survey, see, for example, Quinn *et al.*[7]). In the current debate of the retirement literature it is generally believed that retirement is a dynamic concept that requires explicit incorporation of features of the social security system and the pension system. Alternatively, reduced form estimation of such models is feasible, but would require flexible specification. This means that extension of equation (1) with a labour supply model and a model for $f_2(S_t)$ would make the analysis cumbersome. Moreover, as we sample from an ongoing process, initial condition problems may arise in the sense that γ may not be orthogonal to the included regressors in X_t. This would be particularly relevant if X_t included aspects of health related behaviour (such as smoking, drinking and exercising) or aspects of working conditions. Of course, lagged health status could be included to capture the effect of this reversed causality, but this would require an extension of the model, making estimation even more complex.

For our purposes, where main interest is in the parameters of equation (1), we do not need a complete model of h_t, $f_1(L_t)$, $f_2(S_t)$ and X_t. Alternatively, a random effect approach could be employed that specifies the dependence of γ on $f_1(L_t)$, $f_2(S_t)$ and X_t, but then identification of the causal effect of $f_1(L_t)$, $f_2(S_t)$ and X_t relies on the assumed dependence of γ on these.

We therefore proceed in a way that treats the individual effects γ as unknown fixed parameters. Effectively this means that we estimate the model conditional on the values of the individual fixed effects that need to be estimated along with the other parameters of the model. This approach requires no assumptions on the dependence structure of the regressors on the right hand side of equation (1) on γ. However, estimation of the health level equation (1) with individual specific constants increases the dimensionality of the problem enormously and we therefore take the first difference of equation (1) to obtain

$$\Delta h_t = \alpha_1 \Delta f_1(L_t) + \alpha_2 \Delta f_2(S_t) + \beta' \Delta X_t + \Delta\varepsilon_t \quad (2)$$

Δ is the first difference operator, i.e., $\Delta h_t = h_t - h_{t-1}$. Owing to the differencing the remaining stochastic variation is $\Delta\varepsilon_t$ which is purged from the permanent component γ. As a consequence, consistent estimation of the parameters of interest does not require the specification of equations describing L_t and S_t and/or their relation to γ. OLS estimates of equation (2) yield consistent estimates of α_1, α_2 and β. This property depends on the assumption that the expected value of the random disturbance in equation (1) conditional on L_t and S_t is constant within the relevant time window. This assumption reflects the idea that decisions regarding career and health are jointly made and show up as upward or downward shifts in individuals' health profiles that will typically be correlated with their labour market histories and previous health shocks. With a three or more wave panel one could explicitly test this assumption of time constancy by, for example, allowing the shift in the health profile to be age dependent.

Though differencing of equation (1) is convenient given the model assumptions, it has to be noted that along with γ all time constant regressors in equation (1) will cancel from the specification in equation (2). This means that estimation of equation (2) alone will not allow us to assess the effect of, for instance, gender and cohort effects in explaining differences in health outcomes. In the next section a procedure is employed that allows us to recover the individual fixed effects from estimates of the difference equation (2) and to relate these to time constant regressors that were omitted from equation (2). We first start with a discussion of some issues concerning the implementation of the model.

MODEL IMPLEMENTATION, ESTIMATION ISSUES AND RESULTS

Implementation

As discussed above, we concentrate on (changes) in general health instead of work related health, using the transformed and untransformed outcomes of the Hopkins Symptom Checklist (HSCL). The HSCL scores are measured on a scale ranging from 0 (very healthy) to 171 (very unhealthy) for the total score and from 1 (very healthy) to 7 (very unhealthy) for the transformed score. For applications in labour supply and retirement models, a work related measure would be more appropriate. The use of the HSCL measure will therefore 'limit' our analysis to changes in general health status instead of changes in work capacity.

The effect of labour market status and/or work history $f_1(L_t)$ will be operationalized by a set of labour market status variables. We will use variables indicating the labour market state at a point in time, say L_t, and a range of variables for the number of months spent in work for a short period preceding the date *t* (let us say 2, 5 and/or 10 years), $Ls_{E,t}$ and the number of months spent during the entire working life in work ($Ll_{E,t}$). Analogously, non-work variables $Ls_{N,t}$ and $Ll_{N,t}$ can be defined. $f_2(S_t)$ will be operationalized by a variable indicating whether or not a health shock has occurred in the preceding 2 year interval. We will omit the function $f_2(.)$ and simply use S_t to denote the variable that indicates whether or not a shock has occurred in the time interval $(t-1,t)$. Implicitly it is assumed that health shocks have a permanent effect on health levels [see equation (1)]. As a consequence, for the difference in health status, Δh_t, only the occurrence of a health shock between the time interval October 1993 and October 1995 are of interest (S_{95}). Of course, alternative specifications, for instance where the effect of the shock on the level of health depends upon the time since the occurrence of a shock, could be exploited. We have access to the 1993 and 1995 waves of the CERRA panel survey. The data provide information on the occurrence of a health shock in the time interval 1993–95, it lacks this information for the period 1991–93. This information is required to estimate models with non-permanent effects of health shocks on the level of health. For that reason, we restrict our attention to permanent effects of health shock on health outcomes.

Estimation issues

Below we present a simple two-stage estimation procedure to obtain estimates from a fixed effect panel data model. The individual effects that were incorporated to account for simultaneity between an individual's labour market history and labour market attachment and its health development are treated as unknown (nuisance) parameters. Completion of both stages will provide us with consistent estimates of the underlying parameters of the health level equation (1).

In the *first stage* the nuisance parameters are eliminated by taking the first differenced health equation (2). The first differenced equation is estimated using least-squares methods. Apart from the individual effects, this also removes all exogenous variables that are constant over time. The use of, notably the transformed, HSCL measure may introduce heteroscedasticity in the error term. Therefore, White heteroscedasticity corrected t-values will be presented in this first step.

In the *second stage* we use first stage estimates from $\Delta h_t = \alpha_1 \Delta f_1(L_t) + \alpha_2 \Delta f_2(S_t) + \beta' \; \Delta X_t + \Delta\varepsilon_t = \Delta V_t \beta + \Delta\varepsilon_t$, to calculate the individual fixed effects from (see, for instance, Hsiao[8]):

$$\hat{\gamma} = \bar{h} - \hat{\beta}\bar{V} \quad (3)$$

Next, these (computed) fixed effects are regressed on a vector of time constant exogenous variables that were excluded in the first stage, i.e. $\hat{\gamma} = Z'\delta + \nu$, where Z' includes cohort variables, gender, life style variables, etc.

A closer look at the second stage regression is required to obtain the correct standard errors of $\hat{\delta}$. Note first that the parameter γ satisfies $\gamma = \bar{h} - \beta\bar{V} - \bar{\varepsilon}$ and that instead we use equation (3) to obtain an estimate $\hat{\gamma}$. So the error term ν of our second stage regression satisfies

$$\nu = u + \bar{V}\,(\beta - \hat{\beta}) + \bar{\varepsilon} \quad (4)$$

with $\bar{\varepsilon} \perp u$. νrepresents three sources of uncertainty that enter the second stage regression. u is the unexplained part of the true model (i.e. from $\gamma = Z'\delta + u$) and the last two terms follow from the difference between γ and its estimate obtained from equation (3). If we take $w = \bar{\varepsilon} + u$, then it follows that Cov(ν) can be written as

$$\text{Cov}(\nu) = \sigma_w^2 I_n + \bar{V}\,\text{Cov}(\hat{\beta})\,\bar{V}' \quad (5)$$

where I_n is the identity matrix with rank equal to the number of observations (n). Define K_Z as the rank of the matrix Z, $\hat{w} = \hat{\gamma} - Z\hat{\delta}$ and $M_Z = I_n - Z(Z'Z)^{-1}Z'$, then it can be shown that a consistent estimate of σ_w^2 can be derived from

$$(n - K_Z)\hat{\sigma}_w^2 = \hat{w}'\hat{w} - \text{trace}\,[\bar{V}\,M_Z\,\bar{V}\,\text{Cov}(\hat{\beta})] \quad (6)$$

Next, equations (5) and (6) can be used to obtain an expression for the variance-covariance matrix of $\hat{\delta}$ [Côv ($\hat{\delta}$)]. More specifically, the standard errors of the second stage estimates can be derived from

$$\text{Côv}(\hat{\delta}) = (Z'Z)^{-1} Z'\,\text{Côv}(\hat{\nu}) Z (Z'Z)^{-1} \quad (7)$$

Results

The model is estimated on a subsample of 2422 heads of household that completed the HSCL questionnaire in both waves of the CERRA data. The results of the fixed effect panel data model are reported in Tables 1 and 2. Both results for the total HSCL score and the value on a seven-point scale that is usually derived from that are reported. Table 1 provides estimates of the parameters of the difference equation and Table 2 of the regression of the fixed effects on a range of time constant regressors. We start with a discussion of the results in Table 1.

The intercept in this equation reflects the linear effect of age on the HSCL score. Age squared in the equations for the HSCL level in 1993 and in 1995 turns up as a linear age effect in first-difference equation. The interaction of gender and age in the levels equation turns up as a gender dummy in the first difference equation. The effect of labour market status on the level of health is captured by the first differences of the corresponding variables. The effect of work (wage earner), for instance, is measured by the difference in the outcomes of this variables for both waves (1995 and 1993). Dummy variables for the other labour market status are defined accordingly. The Unemployed are the reference group. Also lagged labour market status variables and the (differences) in the number of months worked in the past 2, 5 and 10 years are included to measure the effect of labour market history.

The results in Table 1 are strikingly different from the results of the simple OLS regression reported above. The estimates for the coefficients of age and gender imply that health deteriorates over the relevant age range for males. For females health deteriorates onwards from age 52. This contrasts with the results from Table A3, where health was found to improve with age. Compared with Table A3, there are large differences in the effect of the labour market variables. The significant negative effects of Work, Self-Employment and Early Retirement in Table A3 suggested a positive effect on the health condition. The results from this table contrast strongly and would suggest that work affects health. Also, the effect of the number of months worked in the last 10 years becomes significant and changes sign, suggesting that a long work history has a negative effect on health. The latter result, however, should be seen in combination with the results from the

Table 1. First stage estimates: linear regression of change of HSCL score

Variable	HSCL on 7-point scale		Total HSCL score	
	Estimate	*t*-value[a]	Estimate	*t*-value[a]
Constant	–0.6245	2.27 (2.35)	–6.0405	2.42 (2.49)
Age in 1993[b]	0.0146	2.83 (2.89)	0.1336	2.84 (2.82)
Dummy female	–0.1052	1.48 (1.46)	–0.8357	1.30 (1.19)
First differences of:				
Dummy partner	–0.2907	1.52 (1.43)	–2.3907	1.38 (1.19)
Dummy work	0.2543	1.87 (1.77)	1.2567	1.02 (0.97)
Dummy disabled	0.2205	1.35 (1.00)	0.3730	0.25 (0.15)
Dummy early retired	0.0845	0.79 (0.79)	0.0904	0.09 (0.10)
Dummy self employed	0.4692	1.79 (1.77)	3.5749	1.51 (1.66)
Dummy work (–2 yrs)	0.0987	0.60 (0.60)	0.3687	0.25 (0.23)
Dummy disabled (–2)	0.3841	1.75 (1.55)	5.6067	2.82 (1.68)
Dummy early ret (–2)	0.3396	2.20 (2.38)	2.0833	1.49 (1.57)
Dummy self empl (–2)	–0.2587	0.75 (0.62)	–1.7471	0.56 (0.48)
Months worked in last:				
2 years	0.0062	0.91 (0.86)	0.0757	1.21 (1.09)
5 years	–0.0101	2.01 (1.97)	–0.1218	2.67 (2.56)
10 years	0.0101	2.72 (2.49)	0.1034	3.06 (2.65)
Negative health shock	0.4040	3.49 (2.97)	3.2101	3.06 (2.67)
Positive health shock	–0.3342	1.28 (1.43)	–2.5756	1.09 (1.35)
R^2 Square	0.0231		0.0229	
F	3.5595		3.5174	

[a]Absolute *t*-values and White heteroscedasticity corrected *t*-values in parentheses.
[b]Age in 1993 in the difference equation can be related to the effect of age squared in the health level equation.

other work history variables (the lagged status variables and the number of months worked in the past two years and past 5 years). The F-statistics for the test of joint significance of the set of labour market variables (11 degrees of freedom) are 3.062 and 2.916 for the untransformed and the transformed HSCL score, respectively. The hypothesis that the 11 labour market variables do not matter is strongly rejected (the p-values are 0.00044 and 0.00079, respectively).

To illustrate the joint effect of the labour market variables, we will present some calculations for different types of working careers below. From a comparison of the results of Tables A3 and 1 it can be deduced that it may be hazardous to ignore the simultaneity between health and labour market status and labour market history variables. Using panel data rather than a cross-section one can take account of that type of endogeneity and also distinguish cohort effects from pure age effects.

Two health shocks variables are included in Table 1 to assess the relative importance of positive and negative health shocks in explaining health changes. From Table 1 it is difficult to assess the relative importance of these effects directly. We therefore confronted the effect of a negative health shock in the equation for the untransformed HSCL score with a change in health due to a pure aging effect. The results on the age variables imply that 2 years of aging for a 55 year old male are equivalent to a deterioration of health of 1.22 on the HSCL score. The coefficient for the negative health shock implies that a negative shock for a 55 year old male is equivalent to the effect of 5.27 years of aging on health. A negative health shock causes large changes in health.

The model discussed in the previous section allows for direct effects of age and labour market variables on health. It is conceivable that these key variables may also indirectly influence health levels through their effect on S_t; the occurrence of a health shock. In that case the total effect of labour market variables on health outcomes consists of a effect through S_t and a direct effect

Table 2. Second stage estimates: regression of fixed effects

Variable	HSCL on 7-point scale		Total HSCL score	
	Estimate	t-value[a]	Estimate	t-value[a]
Constant	–7.720	2.92	–88.285	3.53
Age in 1993	0.766	4.66	7.429	4.90
Age squared	–0.008	5.22	–0.076	5.37
Dummy female	–0.185	0.24	–0.647	0.09
Dummy smokes	–0.041	0.75	–0.232	0.44
Dummy drinks	–0.298	3.68	–3.516	4.55
Dummy drinks >3	0.025	0.35	0.312	0.45
Dummy exercise	–0.153	2.75	–1.637	3.08
Profession:				
White collar/high	0.055	0.26	1.219	0.63
White collar/low	0.106	0.47	2.453	1.18
Blue collar/high	0.139	0.56	2.023	0.87
Blue collar/low	–0.016	0.08	0.553	0.29
Education:				
Low/general	0.242	2.68	2.253	2.61
Interm/general	–0.007	0.07	–0.347	0.39
Interm/vocat	0.124	1.39	0.904	1.07
High/general	0.124	0.94	1.132	0.90
High/vocat	–0.007	0.08	0.097	0.11
University	0.100	0.76	0.692	0.55
Age[a]female	0.059	1.55	0.452	1.30
% of life worked (Lab. M. attachment)	–0.018	4.42	–0.157	4.85
R^2 Square	0.5639		0.4948	
F	163.4967		123.8098	

[a]Absolute t-values from equation (7).

on health. We therefore also estimated a (reduced form) model for S_t. The results show some ambiguity. We report on this in Appendix B.

Table 2 reports results of the second stage regression of the fixed effects. Personal characteristics are included such as age, gender and the presence of a partner in the household. Note that age measures the effect of the birth cohort. Different birth cohort effects are allowed for males and for females. Furthermore, we included a range of variables for educational level and occupational level. The dummy variables smoking, drinking and exercising are included to measure life style effects. The percentage of months worked (employed or self-employed) after finishing school and prior to the 1993 interview is included as instrument for labour market attachment. There appear to be strong cohort effects. We furthermore find strong effects of the life style variables (smoking, drinking and exercising) and the labour market attachment variable. Note that the labour attachment variable reflects the aspects of the endogeneity found in the simple least-squares regression discussed earlier (Table A3 in Appendix A). In that regression, work had a health improving effect. The fixed effect panel data model separates this effect out into a true work effect (Table 1) and a spurious work effect (Table 2). Finally, we find little effect for education and occupation in Table 2.

The R^2 values of the first stage estimates are low as can be expected from differenced equations. From the residuals of that equation we can compute pseudo R^2 measures for the explained variation of the health level equations. For the transformed HSCL score these are 0.69 and 0.68 for 1993 and 1995, respectively. The pseudo R^2 for the untransformed scores are 0.71 and 0.72 for 1993 and 1995, respectively. In Figs A1–A4 in Appendix A, predicted health levels are compared with actual values for both 1993 and 1995.

As was noted earlier, approximately 26% of the respondents sampled in the first wave of our survey (1993) were for unknown reasons not present in the second wave held in 1995. Our analyses are based on respondents who participated in both waves of the survey. The results may be biased if attrition is non random to the (health) variable of interest. We therefore performed some simple tests on the (non) randomness of attrition in our survey. It can be concluded from these tests that the hypothesis of random attrition cannot be rejected. We report on this in Appendix C.

The results from Tables 1 and 2 provide estimates of the effect of age and labour market status and labour market history on (changes in) health levels. From notably the results in Table 1, it is difficult to assess the effect of age on health as the first difference estimates need to be transformed first to make the results interpretable. For that purpose we used Tables 1 and 2 to perform two types of calculations with the model: first a calculation of age, gender and health profiles, then profiles of health for different labour market states and levels of work experience.

Calculations with the model: age and gender profiles for different cohorts

From the estimates in Tables 1 and 2, we can derive the following relationship between the untransformed HSCL score and age, sex and birth cohort (standard errors in parentheses):

$$h = \gamma - \underset{(1.27)}{3.087}\,\text{Age} + \underset{(0.012)}{0.033}\,\text{Age}^2 - \underset{(0.322)}{0.418}\,\text{Age*female}$$

$$\gamma = \underset{(23.05)}{-54.30} + \underset{(1.19)}{6.698}\,\text{Birthyr} - \underset{(0.015)}{0.076}\,\text{Birthyr}^2 + \underset{(33.36)}{41.40}\,\text{Female} - \underset{(0.128)}{0.452}\,\text{Female*Birthyr}$$

For the transformed HSCL scores a similar relationship can be derived. The remaining regressors are taken as fixed at values for typical respondents. This implies that in Figs 1 and 2 attention should focus on the pattern of age over time and the distance between different lines in the figures rather than focusing on the level of health at specific ages.

The figures depict health profiles over age for different cohorts of males and females. The dotted line represents age health profiles for different cohorts of females and the solid lines is for males. The figures depicts large differences in health levels for different cohorts. Cohort effects are measured by a quadratic function with a 'top' of 1942 for males and 1944 for females. As a result, we find, on average, worse health levels for male and female cohorts born during the Second World War. This may be interpreted as indication that differences in the environment and nutrition intake in early childhood have long-term effects on health outcomes. For females, the 1950 cohort (the youngest cohort depicted in the figure) is the

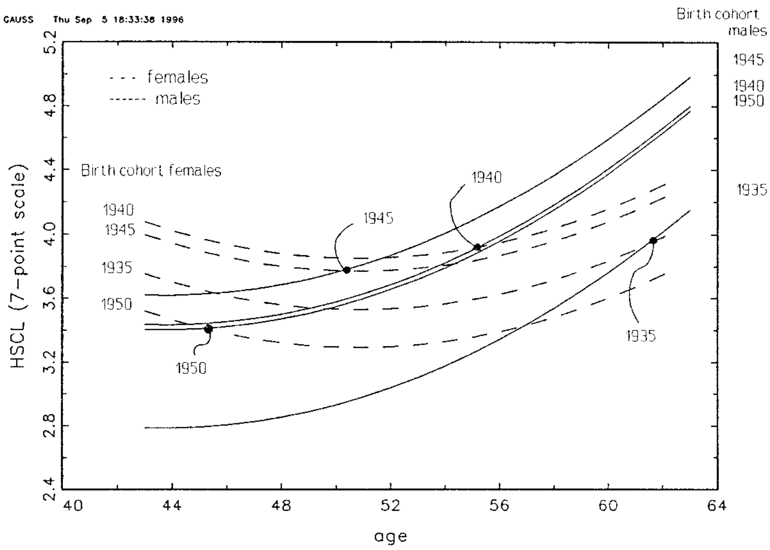

Figure 1. Age–health profiles for males and females from different birth cohorts (HSCL on seven-point scale; low = healthy).

most healthy. For males, the 1935 cohort (the oldest cohorts depicted in the figure) appears to be the most healthy. This effect for males is surprising. It may be the case that the quadratic specifications of age and cohort are too restrictive. We estimated alternative models with spline functions for age and cohorts. This did not alter the results. An alternative explanation is an effect that we denote as a 'survivor' effect. It could be the case that only respondents in good health remain as the population ages. Hence the oldest cohort may consist of a relatively homogeneous group of healthy survivors, whereas the subgroup of younger cohorts is more heterogeneous in the sense that they still consist of both healthy and less healthy individuals.

For males health deteriorates monotonically with age. For females, health improves up to roughly age 50 and deteriorates thereafter. This may reflect that health deteriorates faster with age at ages beyond the menopausal period. Alternatively, it may be the case that this pattern is due to the small number of females in the left tail of the age distribution. With respect to this, it should be noted that the sample only includes heads of households and that only 18% of the sample consists of females. At the start of the age range that we consider (43 years), females are less healthy than their male counterparts. The health deterioration rates of males are, however, larger than those of females, leading to better health conditions for females at more advanced ages. This is in line with results from published life tables. As a last remark on these figures, the points at which each cohort of males and females intersect seem to come at earlier ages for the youngest cohorts. This may imply that females become, relative to men, more healthy over time.

Calculations with model: the effect of labour market status and labour market history

Several variables relating to an individual's labour market history are included in the specification. To see how these effects operate on the transformed HSCL score health profiles are reported in Table A4 in Appendix A. In Fig. 3 we depict the calculations for the total HSCL scores. The table and the figure make a *ceteris paribus* comparison of the age-health profile of three different types of individuals. All three are male and have worked continuously until the age of 44. Type I continues to work until age 65. The type II individual continues to work and applies for an early retirement scheme at the age of 55. Type III immediately loses his job and stays out of work until he is 65. Comparison of the first and the third types of individual in Table A4 shows that working speeds up the process of health deterioration. The effect of retiring is also marked. The early retiree quickly gains on the worker and the

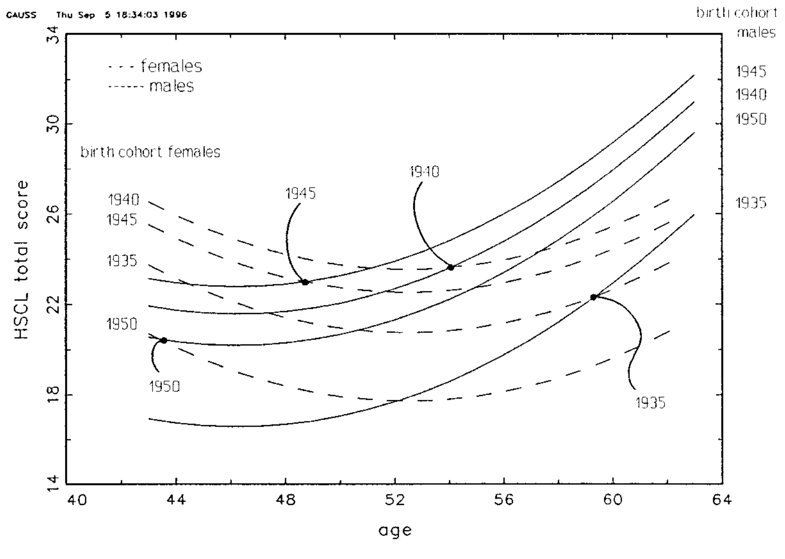

Figure 2. Age–health profiles for males and females from different birth cohorts (HSCL total score; low = healthy).

gap between him and the type III person decreases over the years.

A similar profile shows up in the estimates based on the total HSCL-score. In Fig. 3 we see that after an initial health improvement the retiree experiences a fall-back to almost the health level of the individual that continued to work. Only after 6 years retirement does his health improve relative to that of the worker and the unemployed. The estimates are based on bi-annual information on the respondents' labour market status. This may account for the abrupt turns in the age–health profile after retirement. With more detailed information, preferably month to month information, one would expect to find a more gradual deviation from the age–health profile of workers, implying that it takes several years before a retiree's health improves relative to what his or her health would have been if he or she had continued to work.

CONCLUSIONS

We have focused on aspects of health changes, the importance of cohort effects, age related health changes and the effect of labour market status and work history on health. We have moreover assessed the relative importance of gradual changes in and sudden shocks to health and the role of work status on the likelihood of experiencing a health shock. For that purpose we constructed a fixed effect panel data model that allows for the endogeneity of labour market behaviour and health. A simple two-stage regression procedure was proposed and applied to two waves of a survey of Dutch elderly. We find that it is important to correct for the endogenous interrelation of health and labour market behaviour in a (behavioural) model for health and that panel data are required to disentangle cohort effects from pure age effects. We find differences in health outcomes for different age cohorts and gender. Second World War cohorts have lower health levels that other cohorts. Health deteriorates with age. Health deterioration rates of males are larger than those of females, causing females to be healthier than males at advanced ages. We furthermore find that work affects health, i.e. health deteriorates with employment and labour market history.

ACKNOWLEDGEMENTS

The authors are members of CERRA (Centre for Economic Research on Retirement and Ageing), Department of Economics, Leiden University. This research was financed by a grant from NESTOR (The Netherlands Institute on Ageing) and the Leiden University Speerpunten programme. We acknowledge the valuable comments of Owen O'Donnel, two anonymous referees and participants at the Fifth European Workshop on Econometrics and Health Economics.

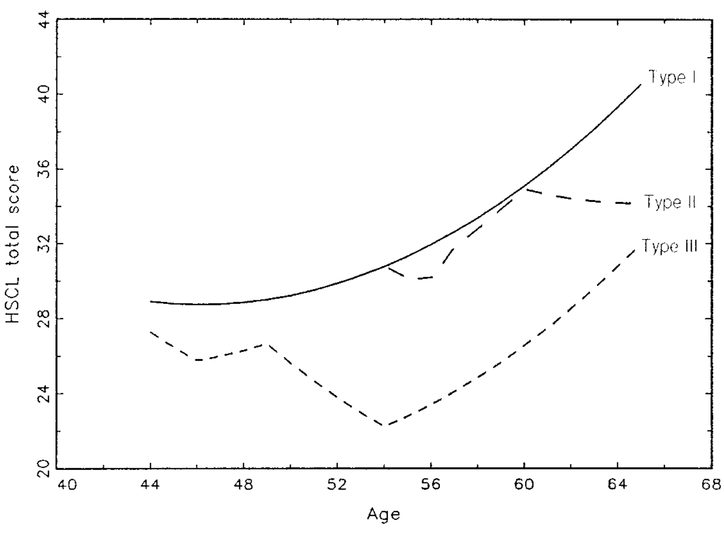

Figure 3. Age–health profiles for three different labour market patterns (HSCL total score; low = healthy).

APPENDIX A

Table A1. Means of main variables in the analyses

Variable	Mean
HSCL[a]	
Total score 1993	13.19
7-point scale 1993	2.79
Total score 1995	13.64
7-point scale 1995	2.89
Age in 1993	55.52
Female	0.18
Living with partner in 1993	0.77
Education:	
Low/general	0.18
Interm/general	0.14
Interm/vocational	0.16
High/general	0.05
High/vocational	0.17
University	0.05
Profession:	
White collar/high	0.41
White collar/low	0.09
Blue collar/high	0.04
Blue collar/low	0.32
Life style variables:	
Smoking	0.39
Drinking	0.84
Exercising	0.45
Labour market history: number of months worked:	
in past 6 months measured in 1993	3.31
in past year measured in 1993	6.76
in past 2 years measured in 1993	14.12
in past 5 years measured in 1993	39.29
in past 10 years measured in 1993	86.87
in past 6 months measured in 1995	2.82
in past year measured in 1995	5.77
in past 2 years measured in 1995	12.10
in past 5 years measured in 1995	34.53
in past 10 years measured in 1995	80.54
Percentage of time working in life	72.39
Labour market status:	
Wage earner in 1991	0.57
Self-employed in 1991	0.07
Early retired in 1991	0.09
Disabled in 1991	0.16
Unemployed in 1991	0.12
Wage earner in 1993	0.47
Self-employed in 1993	0.07
Early retired in 1993	0.16
Disabled in 1993	0.16
Unemployed in 1993	0.14
Wage earner in 1995	0.38
Self-employed in 1995	0.06
Early retired in 1995	0.20
Disabled in 1995	0.15
Unemployed in 1995	0.20

Table A1. *Continued*

Variable	Mean
Shocks in health status	
Shocks that have a negative impact on health	0.06
All shocks (positive and negative)	0.07

[a]Total score of the Hopkins Symptom Checklist (HSCL) ranges from 0 to 171; low scores are associated with good health, high scores with bad health. The transformed HSCL score is a seven-point scale ranging from 1 (very good) to 7 (very bad).

Table A2. Means of total HSCL scores by labour market status and age group in 1993

	Age[a]					
	1	2	3	4	All	1995–93
Employed	9.61	10.63	10.38	7.45	10.08	0.77
Early Retired	–	8.00	7.40	8.91	8.28	2.55
Disabled	28.65	30.91	24.85	23.23	25.95	0.28
Unemployed	18.80	19.91	17.40	13.41	16.67	0.25
Self-employed	8.15	7.94	11.90	8.88	9.34	0.83

[a]1, age ∈ <→,[50]; 2, age ∈ [51,55]; 3, age ∈ [56,60]; 4, age ∈ [61,→>.

Table A4. Simulation of effect of labour market status for a male who has worked until the age of 44 and follows one of three scenarios over the next 22 years

Age	Type I: always employed	Type II: retires at 55	Type III: unemployed from 44 onwards
44	4.5396	4.5396	4.1866
45	4.5457	4.5457	4.1183
46	4.5592	4.5592	4.0574
47	4.5799	4.5799	4.0781
48	4.6080	4.6080	4.1062
49	4.6435	4.6435	4.1417
50	4.6862	4.6862	4.0632
51	4.7363	4.7363	3.9921
52	4.7936	4.7936	3.9282
53	4.8583	4.8583	3.8717
54	4.9304	4.9304	3.8226
55	5.0097	4.8399	3.9019
56	5.0964	4.8522	3.9886
57	5.1903	5.1126	4.0825
58	5.2916	5.2139	4.1838
59	5.4003	5.3226	4.2925
60	5.5162	5.4385	4.4084
61	5.6395	5.4406	4.5317
62	5.7700	5.4499	4.6622
63	5.9079	5.4666	4.8001
64	6.0532	5.4907	4.9454
65	6.2057	5.5220	5.0979

Table A3. Excerpts from OLS regression of the level of HSCL scores, transformed and total (untransformed) score in 1993[a]

	HSCL	
	7-point scale	Total score
Constant	–3.780 (1.4)	–56.158 (2.2)
Employed (wage earner) in 1993	–0.714 (3.3)	–6.444 (3.0)
Self-employed in 1993	–1.130 (2.8)	–8.919 (2.3)
Early retired in 1993	–0.741 (4.4)	–5.909 (3.6)
Disabled in 1993	0.680 (3.0)	5.387 (2.4)
Employed (wage earner) in 1991	–0.031 (0.1)	–1.096 (0.5)
Self-employed in 1991	0.300 (0.8)	1.738 (0.5)
Early retired in 1991	0.454 (2.2)	3.313 (1.7)
Disabled in 1991	0.540 (2.2)	6.287 (2.7)
Number of months worked in last 2 years	0.019 (1.5)	0.116 (0.9)
Number of months worked in last 5 years	0.004 (0.6)	0.094 (1.5)
Number of months worked in last 5 years	–0.004 (1.8)	–0.053 (2.4)
Age	0.285 (2.9)	3.041 (3.2)
Age squared	–0.0027 (2.9)	–0.029 (3.3)
R^2	0.18	0.19
F	22.60	23.02

[a]The regession contains a range of other controls such as occupation, education, gender, marital status and life style variables such as smoking, drinking and exercising.

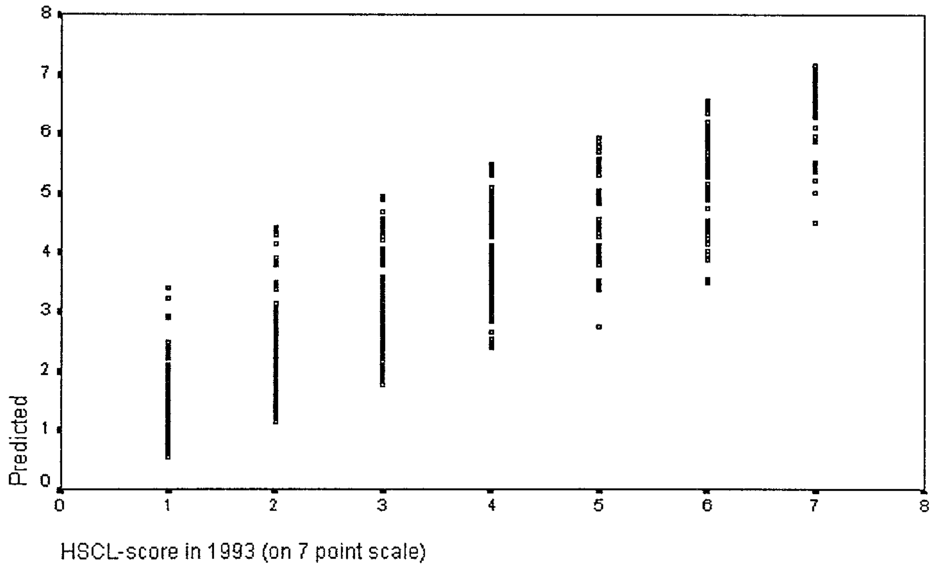

Figure A1. Predicted versus actual 1993 HSCL score (seven-point scale).

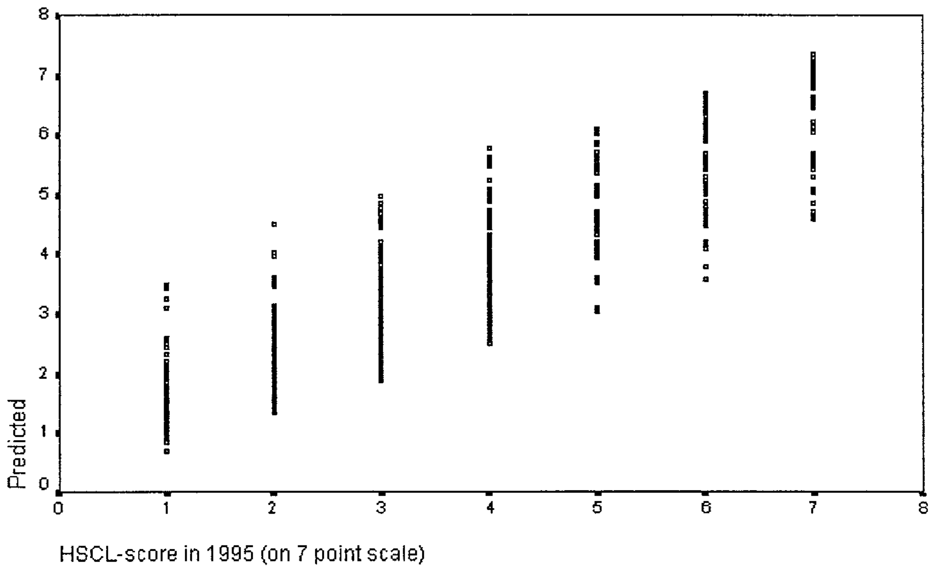

Figure A2. Predicted versus actual 1995 HSCL score (seven-point scale).

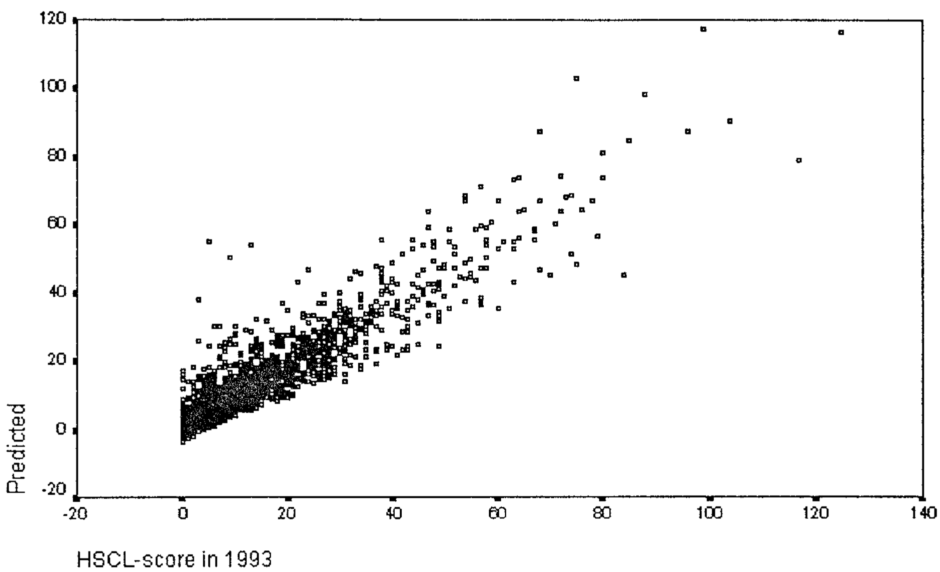

Figure A3. Predicted versus actual 1993 HSCL score (total value).

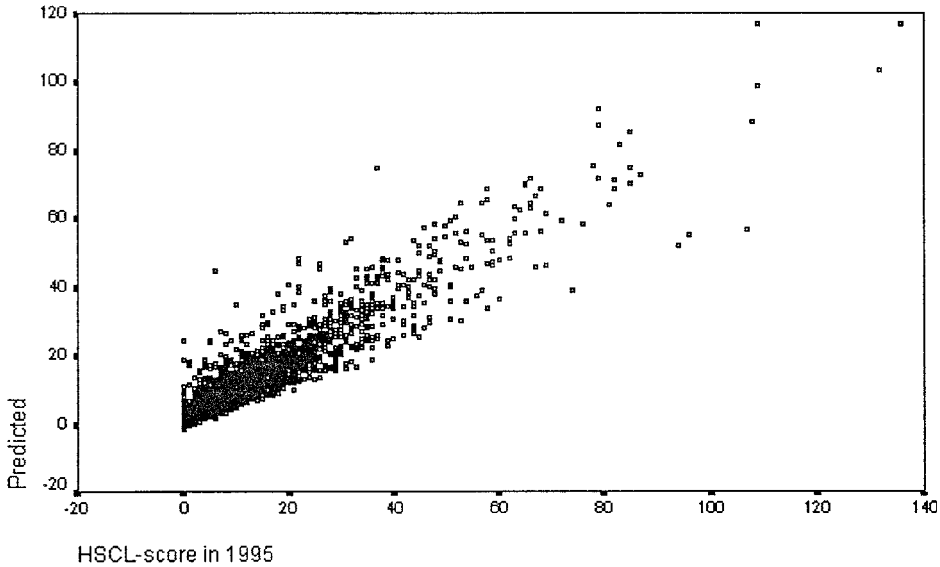

Figure A4. Predicted versus actual 1995 HSCL score (total value).

APPENDIX B

A model for the occurrence of a health shock S_1

In modelling S_t, we take the traditional approach and assume that the indicators S_t are governed by an underlying latent construct S^*_t and assume that S_t equals 1 if and only if the underlying latent construct exceeds a certain threshold (normalized to zero) and zero otherwise. S^*_t may be specified as

$$S^*_t = \delta_0 + \delta_1' W_t + \xi_t \qquad \text{(B1)}$$

The vector W may contain a range of variables of which labour market status variables and age may be the most prominent ones. Some comments remain before we turn to the estimation results of this probit model. First, health level cannot be included in the model as the model in the text implies that health shocks and health levels may be related through the unobserved component γ. Second, inclusion of labour market variables in W requires an exogeneity assumption. S_t is observed once (between 1993 and 1995, so it is not possible to apply a procedure similar to that for the health level equation). Extending the model with a model for labour market behaviour would be a solution, but we feel that this is beyond the scope of this paper. Finally, as we observe S_t only once, is impossible to distinguish pure age effects from cohort effects. We now briefly turn to the estimation results reported in Table B1.

The table reports results on the probability of experiencing a health shock in the 2 year time interval (1993–95). The included regressors are taken at their 1993 value. The probability of experiencing a health shock increases with age, though (again) it has to be noted that the age variable will also capture an age cohort effect. Individuals who were (previously) employed in the construction sector on average have a higher probability of experiencing a sudden change in their health condition. The coefficient of current and lagged labour market status are approximately equal but opposite in sign. This means that those (still) at work on average have lower health shock probabilities than those out of work. Although this indeed may reflect differences in the way in which health changes for workers and non-workers, this may also signal simultaneity problems in the regression. It may be the case that individuals are still at work in 1993 because they have on average lower risk of experiencing health shocks. Since work status variables are taken at the start of the time interval (i.e. prior to the occurrence of a health shock), the alleged simultaneity must run through unobservables associated with work status and health shock variables. As

Table B1. Probit estimates of the occurrence of health shocks (all shocks and only negative shocks

	All shocks		Negative shocks	
	Estimate	*t*-value	Estimate	*t*-value
Constant	–2.849	–5.67	–2.622	–5.04
Female	0.202	1.45	0.170	1.18
Age in 1993	0.020	2.42	0.017	1.99
Partner	0.122	0.97	0.057	0.44
Working in 1993	–0.657	–2.19	–0.416	–1.24
Working in 1991	0.547	2.40	0.624	2.67
Months (self-)employed at 1993 in:				
Previous 6 months	0.040	0.29	–0.090	–0.60
Previous year	0.101	1.04	0.184	1.75
Previous 2 years	–0.044	–1.33	–0.073	–2.03
Previous 5 years	–0.004	–0.42	0.002	0.20
Previous 10 years	–0.0001	–0.02	–0.002	–0.63
Education:				
Low/general	0.203	1.87	0.127	1.12
Interm/general	–0.074	–0.57	–0.123	–0.92
Interm/vocational	0.041	0.35	–0.067	–0.55
High/general	0.177	1.07	0.156	0.92
High/vocational	–0.061	–0.52	–0.104	–0.85
University	–0.158	–0.84	–0.185	–0.96
Sector:				
Industry	–0.027	–0.29	–0.063	–0.64
Construction	0.266	2.09	0.198	1.46
Transport	–0.343	–1.68	–0.275	–1.34
Banking	0.036	0.26	0.026	0.19

our data on the health shock variables is limited, we did not pursue this model any further.

APPENDIX C

Evaluating the importance of attrition

In the text, it was noted that a substantive fraction (26%) of the respondents sampled in the first wave (1993) were for unknown reasons not present in the second wave held in 1995. The estimates presented in this paper are based on a balanced panel. These results may be biased if the attrition is endogenously related to the (health) variable of interest. It is conceivable that individuals with intrinsically bad health are not able or are not willing to participate in a second survey. In that case, the sample survivors in the second wave will be relatively healthy. It could also be the case that individuals who make transitions in the labour market change place of residence and as a consequence it will be more difficult for the survey bureau running the survey to trace these individuals. In that case attrition may cause relatively immobile individuals to be over-represented in the second wave of the survey. Since labour market behaviour and health are endogenously related, attrition will also affect the distribution of health non-randomly. Non-random attrition requires the selection process to be modelled jointly with the process of the variable of interest.

To assess whether attrition is non-random to our variable of interest, we perform a simple formal test. If attrition is non-random to the distribution of health, then one would expect that the unobservables in a reduced form equation for the presence of attrition to be related with the unobservables of the reduced form health level equations of 1993 and 1995. Since the distribution of health in 1993 is unaffected by the attrition process, one can test for association between the

Table C1. Test on non-random attrition: OLS regressions of health status in first wave with a dummy for no attrition in future wave

	HSCL 7-point scale		HSCL total score	
	Estimate	*t*-value	Estimate	*t*-value
Constant	–4.700	–2.14	–62.901	–2.88
No attrition in wave II	–0.013	–0.24	0.117	0.21
Female	0.359	3.76	3.600	3.79
Age	0.282	3.45	2.886	3.55
Age squared	–0.026	–3.38	–0.0271	–3.51
Partner	–0.500	5.90	–4.974	–5.90
Education:				
Low/general	0.276	3.56	2.671	3.48
Interm./general	–0.122	–1.42	–1.460	–1.72
Interm./vocational	–0.103	–1.28	–0.942	–1.17
High/general	–0.198	–1.55	–2.004	–1.58
High/vocational	–0.159	–1.86	–1.500	–1.76
University	–0.241	–1.97	–2.357	–1.94
Profession:				
White collar/high	0.193	2.49	1.837	2.37
White collar/low	0.065	0.61	1.291	1.21
Blue collar/high	0.255	1.77	1.766	1.23
Blue collar/low	0.329	4.08	3.344	4.17
R^2	0.06		0.06	

health process in 1993 and the attrition process using a very simple procedure. A dummy variable for future attrition can be included in the (reduced form) health equation. In case sample non-respondents are individuals with intrinsic low (high) values of health, then the coefficient of the dummy for future attrition will be negative (positive). The results of this simple test are reported in Table C1. From this table it can be concluded that the hypothesis of random attrition can not be rejected. We also performed a similar test in which we include the health level in 1993 as a regressor in a probit equation explaining whether individuals participate in the second wave. This test confirmed the results of the previous test.

REFERENCES

1. Bazzoli, G. The early retirement decision: new empirical evidence on the influence of health. *Journal of Human Resources* 1985; **20**: 214–234.
2. Anderson, K. H. and Burkhauser, R. V. The retirement-health nexus: a new measure of an old puzzle. *Journal of Human Resources* 1985; **20**: 315–330.
3. Stern, S. Measuring the effect of disability on labour force participation. *Journal of Human Resources* 1989; **24**: 361–395.
4. Bound, J. Self-reported versus objective measures of health in retirement models. *Journal of Human Resources* 1991; **26**: 107–137.
5. Kerkhofs, M. J. M. and Lindeboom, M. Subjective health measures and state dependent reporting errors. *Health Economics* 1995; **4**: 221–235.
6. Fuchs, V. R. Time preference and health: an exploratory study. In: Fuchs, V. R. (ed.) *Economic Aspects of Health*, pp. 93–120. Chicago: University of Chicago Press, 1982.
7. Quinn, J. F., Burkhauser, R. V. and Meyers, *Passing the Torch: the Influence of Economic Incentives on Work and Retirement*. W.E. Upjohn Institute for Employment Research, 1990.
8. Hsiao, C. *Analysis of Panel Data*. Econometric Society Monographs. Cambridge: Cambridge University Press, 1986.