JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **24**: 1024–1046 (2009)
Published online 14 April 2009 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/jae.1077

HEALTH AND WORK OF THE ELDERLY: SUBJECTIVE HEALTH MEASURES, REPORTING ERRORS AND ENDOGENEITY IN THE RELATIONSHIP BETWEEN HEALTH AND WORK

MAARTEN LINDEBOOM[a]* AND MARCEL KERKHOFS[b]

[a] *VU University, Amsterdam, The Netherlands; HEB, Bergen, Norway; Tinbergen Institute, Amsterdam, The Netherlands; IZA, Bonn, Germany; and Netspar, Tilburg, The Netherlands*
[b] *OSA Institute for Labor Studies, Tilburg University, The Netherlands*

SUMMARY

This paper explores the interrelation between health and work decisions of older workers. For this, two issues are of relevance. Firstly, health and work may be endogenously related because of direct causal effects of health on work and vice versa, and because of unobservables that may affect both observed health and work outcomes. Secondly, social surveys usually contain self-assessed health measures and research indicates that these may be subject to endogenous, state-dependent reporting bias. A solution to the 'Health and Retirement Nexus' therefore requires an integrated model for work decisions, health production and health reporting mechanisms. We formulate such a model and estimate it on a longitudinal dataset of older Dutch males. Copyright © 2009 John Wiley & Sons, Ltd.

1. INTRODUCTION

This paper addresses the relation between health and labor supply decisions of older workers. For this, two issues are particularly relevant. Firstly, the relationship between health and work is complex, encompassing direct causal effects of health on work and vice versa, and also correlations reflecting the impact of confounding variables, some of which may be unobservable. Secondly, in addition to this 'classical' simultaneity issue, further endogeneity problems may arise by the way in which the relevant health status is observed in social surveys. These usually contain direct self-assessments of the extent to which health limits the ability to do paid work and various indicators of an individual's more general health status, such as the presence of chronic health conditions or checklists of difficulties performing a range of daily activities. In a model of labor force participation or retirement decisions, a work-related health measure may be most appropriate and the use of more general health indicators could lead to biased parameter estimates, due to measurement error. On the other hand, self-reports on health-related limitations to perform work may differ according to the labor market status of the respondent. A person receiving disability pay may rationalize inactivity by exaggerating his or her health problems. It may be socially more accepted to report health problems as the main determinant for inactivity, and for some social security benefits, notably disability insurance benefits, eligibility conditions are contingent upon

* Correspondence to: Maarten Lindeboom, Department of Economics, VU University, Amsterdam, de Boelelaan 1105, 1081 HV Netherlands. E-mail: mlindeboom@econ.vu.nl

bad health (Bound, 1991). Individuals without a paid job might therefore be inclined to overstate health problems. This 'state-dependent' reporting, also known as 'justification', implies that when these measures are used in empirical models one is likely to overestimate the effect of health and underestimate the effect of other variables, such as economic incentives. The relevance of this type of reporting bias will differ across countries as it is likely to depend on the accessibility and generosity of the disability insurance (DI) programs.

Most European countries have social security systems that discourage continued work of older workers (see, for instance, the country studies by Gruber and Wise, 1998, and references therein). The Netherlands used to be an extreme case. Since the mid 1970s, labor force participation rates of older (55+) males have dropped 40% points to a level of about 40% in the early 1990s. Employer-provided early retirement (ER) schemes allowed for retirement at the age of 60, or sometimes even earlier. In addition to these schemes, unemployment insurance schemes (UI) and DI schemes had been used as an alternative route for retirement (see, for instance, Aarts and de Jong, 1992). In the second half of the 1990s ER schemes were gradually made less attractive, and more recently the DI system has been reformed, leading to a rise in the participation rates in the 55+ age range. Strong incentive effects of DI schemes have also been found for a number of other countries. Bound and Burkhauser (2000) noted that 'the prevalence of disability transfer recipients per worker has increased at all working ages over the last quarter of the century in the United States and in the Netherlands, Sweden and Germany. This coincides with an increase in both access to and the generosity of publicly provided social insurance and social welfare programs targeted at people with disabilities in the industrialized world.' This implies that in many countries the stock of DI recipients may not only consist of individuals who are in poor health, which has important consequences for social security policy and for applied research in labor and health economics.

The above implies that an analysis of the interrelation between health and work requires a joint model for work decisions, health *and* health reporting. The contribution of this paper is to specify a model that addresses these methodological problems simultaneously. We estimate it on Dutch longitudinal data. Under the assumptions in the model, it generates consistent estimates of (i) the effect of health and financial incentives on work, (ii) the effect of work (history) on health and (iii) the size of the reporting errors in the self-reported health measure that is used. In addition, we provide a procedure for eliminating state-dependent reporting errors from subjective health data, which may be relevant for a wider range of applications.

The remainder of the article is organized as follows. Section 2 presents the general outline of the model for health, work decisions and health reporting. Section 3 discusses the data that are used in the empirical application. This application and its results are discussed in Section 4 and Section 5 concludes.

2. A MODEL FOR HEALTH, WORK AND HEALTH REPORTING

In order to study the interrelation between health and work at older ages empirically, we need to specify a model that addresses the methodological problems mentioned above. More specifically, the model should describe how work and work history affect health and how health and financial incentives affect the labor force participation decision of older workers. In addition, the model should account for systematic, particularly state-dependent, differences in how people report on the

DOI: 10.1002/jae

extent to which their ability to perform work is limited by health problems. This section presents the general structure of the model and its empirical implementation.

2.1. Health Variables

Most social surveys contain a self-assessed indicator of the extent to which health problems limit the ability to work. Such measures clearly provide the most relevant information for labor supply and retirement models but, as argued in the Introduction, they may be biased towards poor health for some groups of respondents that are out of work. *Justification bias* or *state-dependent reporting* may make such indicators endogenous in the context of labor supply or retirement models and may lead to an upward bias of the estimated effect of health.

Most surveys also contain other health measures that are less likely to be affected by justification bias, but that do not specifically relate to work. Examples are the prevalence of certain conditions, test scores based on checklists of symptoms and self-ratings of health in general. The drawback of these measures is that they relate to an individual's *general* health status, rather than to the ability to perform paid work. As such they are typically noisy indicators of the relevant work-related health status and application in labor supply and retirement models will generally lead to a downward bias of the estimated effect of health. Bound (1991) provides an extensive discussion of the biases involved in the use of different kinds of health measures in retirement models.

2.2. State-Dependent Reporting

The general idea of combining the information from a self-assessed and potentially endogenous work-related health measure and a more general health measure, which is less sensitive to state-dependent reporting, is to use the second measure to correct for systematic response bias in the work-related health measure. There is evidence for justification bias or state-dependent reporting in the self-reports if the difference between the measures shows systematic variation across labor market states. Assuming that the more general health measure is not affected by state-dependent reporting, it can be used as a benchmark to assess what the work-related measure would be without justification bias. This 'cleansed' health measure could then be used in a model for labor supply or retirement decisions. The advantage of explicitly modeling the response mechanism is not only that it offers the opportunity to use the valuable information from the work limitations question for the analysis of the labor supply decision, but also it provides information about the extent of the bias in the work-related health measure.

The comparison between the two health measures is made by modeling the response mechanism explicitly. To fix ideas, let H be an ordinal measure based on a survey question like 'Does your health limit you in the kind and amount of work that you can do?' Typically, the answer will be one out of K ordered response categories. For individual i at time t, the response to the work limitations question can then be described by the latent index model:

$$H_{it} = k \Leftrightarrow c_{k-1}(S_{it}) < H^*_{it} < c_k(S_{it}), \quad \text{for } k = 1, .., K \tag{1}$$

where S denotes the labor market status and c the thresholds that separate the ranges of the latent index H^* that correspond to each of the response categories.[1] Typically, S will also be a

[1] Where $c_0 = -\infty$ and $c_K = \infty$. See Terza (1985) for a discussion of generalized ordered response models.

DOI: 10.1002/jae

categorical variable denoting the possible labor market states. H is affected by labor market status S for two reasons: by the direct (causal) impact of the labor market status on health and by how responses are affected by justification bias. In the context of the latent index specification (1), the first effect is how S affects the latent health index H^* and the second effect is captured by the dependence of the response thresholds c on S. Reporting behavior may also depend on other individual characteristics and therefore one could specify the reporting thresholds as functions of S and X. Coherency of this type of ordered response model requires that the thresholds do not cross, i.e., $c_k(.) < c_{k-1}(.)$. Increasing the number of variables in the specification of the threshold increases the risk of incoherency. Moreover, for our present purpose, we only need to distinguish between the direct causal effect and any justification bias related to S. Here, we will therefore only focus on differential reporting with respect to labor market status S (see also Section 2.5).

We model the latent health index H^* using a more general health measure (H^o) and a set of additional control variables X^*. For identification of the justification bias caused by S we assume that H^o is not affected by justification bias and additionally we have to assume that the direct causal impact of the labor market status on the latent health index H^* is sufficiently captured by its impact on H^o.

Assumption 1: Conditional on H^o, a set of exogenous variables X^* and an individual effect η^*, the labor market status (S) has no additional effect on H^*. Thus pdf $(H^*|H^o, S, X^*, \eta^*) \equiv$ pdf $(H^*|H^o, X^*, \eta^*)$.

This conditional independence assumption implies that in empirical applications we can compare responses to the question about limitations (H) across labor market states, after controlling for exogenous characteristics (X^* and η^*) and another observed health measure that is believed not to be affected by justification bias (H^o).[2] After controlling for H^o, X^* and η^*, any remaining variation of H across labor market states can then be interpreted as justification bias.

In the empirical application (Section 4) we will use the score of the Hopkins Symptoms Checklist (HSCL) for H^o. This is a list of 57 questions that form a validated test of an individual's general health status, which is used in the medical sciences to assess the psychoneurotic and somatic pathology of patients (see Derogatis *et al*., 1974). The HSCL is known to have an excellent rate of external consistency, meaning that the test results are highly correlated with objective medical reports on the patients' health status. In the survey, the complete checklist is part of the self-administered questionnaire (no interviewer). Furthermore, it consists of items related to a wide range of specific somatic and mental health issues and in the questions there is no reference to work or benefits. We therefore think it is reasonable to assume that this measure is not sensitive to justification bias and sufficiently versatile to support Assumption 1.[3] The health limitations variable H is based on the question 'Does your health limit you in the kind and the amount of work that you can do?' The response categories are that health (1) 'causes no problem', (2) 'causes some problems', (3) 'causes severe difficulties' and (4) 'makes it impossible to work', so $K = 4$ in the

[2] H^o does not need to be observed without error. What is required is that the measurement error is not related to the labor market status of the respondent.

[3] At least it would seem reasonable to assume that the HSCL score is less likely to be affected by justification bias than the direct measure of health-related problems regarding paid work (H). If the HSCL score also suffers from justification error, the results show by how much justification error is stronger for the direct health indicator H than for the HSCL score. In that case, the effect of justification bias will be mitigated, though not fully eliminated.

DOI: 10.1002/jae

empirical application. For the labor market status S we distinguish four states: work, unemployed (UI), disabled (DI) and early retired (ER).

The relationship between H^* and H^o is specified as a linear regression model, which by Assumption 1 is now also satisfied conditional on S:

$$H_{it}^* = \alpha \cdot H_{it}^o + X_{it}^* \beta^* + \eta_i^* + e_{it}^* \tag{2}$$

where η^* is an unobserved individual component and e^* is an i.i.d. error term. After substitution of (2) in (1) we get

$$H_{it} = k \Leftrightarrow c_{k-1}(S_{it}) < \alpha \cdot H_{it}^o + X_{it}^* \beta^* + \eta_i^* + e_{it}^* < c_k(S_{it}), \quad \text{for } k = 1, .., K \tag{1$'$}$$

In the empirical application we will assume that the random errors e^* are independently distributed normal random variables with mean 0 and variance 1. The reporting thresholds are defined as separate constants for each labor market state, i.e.

$$c_k(S) = \psi_k^S, k = 1, .., K \text{ and } -\infty = \psi_0^S < \psi_1^S < \psi_2^S < \ldots < \psi_K^S = \infty, \text{ for } S = \text{UI, DI, ER, Work}$$

Under Assumption 1 these sets of state-specific threshold parameters would be identical in the absence of justification bias. Any differences describe the impact of state dependent reporting. In comparing the values of the thresholds across labor market states, we will use the workers as the reference group. This choice is based on the assumption that workers have the least incentive to overstate health problems (see, for example, McGarry, 2004; Kreider, 1999; Kerkhofs and Lindeboom, 1995).

In retirement models the effect of health is usually taken into account by including a dichotomous health indicator D that has the value 1 if the respondent reports having health problems that make it difficult or impossible to work, $(H > 2)$and the value 0 otherwise ($H = 1$ or 2). The expected value of this dummy variable can be computed from (1′) as

$$\mathrm{E}(D_{it}^S) = \Pr(H_{it}^* > c_2^S | H_{it}^o, X_{it}^*, \eta_i^*) = 1 - \Phi(\psi_2^S - \alpha \cdot H_{it}^o - X_{it}^* \beta - \eta_i^*) \tag{3}$$

If for some group of respondents the value of ψ_2 is below that of workers, the effect of justification turns up as a higher expected value of the health dummy D, leading to a biased estimate of the corresponding parameter in the retirement model. For non-working respondents (S = UI, DI, ER) we can use the workers' threshold ψ_2^{Work} to compute $\mathrm{E}(D^{\text{Work}})$, the expected value of the self-reported health dummy if the respondent would have been working. The difference between $\mathrm{E}(D^S)$ and $\mathrm{E}(D^{\text{Work}})$ can be seen as a measure for justification bias in this variable. Following this line of reasoning, $\mathrm{E}(D^{\text{Work}})$ can be used as instrument for the dummy variable D in labor supply or retirement models in order to account for state-dependent reporting of health problems.

2.3. Labor Supply and Retirement

As we are mainly interested in work and retirement from the labor force, we focus on transitions from work to unemployment, disability or (early) retirement. This is done by modeling the

DOI: 10.1002/jae

probability of being in either one of these states or having a paid job at time t, conditional on having a paid job at time $t-1$. This probability is denoted as

$$\Pr(S_{it} = j | S_{it-1} = \text{Work},\ Z_{it}, \mathrm{E}(D_{it}^{\text{Work}}), \xi_i^j) = F_j(Z_{it}, \mathrm{E}(D_{it}^{\text{Work}}), \xi_i^j) \quad (4)$$

where j stands for one of the possible labor market states, Work, UI, DI, and ER. As suggested above, we use the counterfactual variable $\mathrm{E}(D^{\text{Work}})$ rather than the indicator D to eliminate the effect of state-dependent reporting on the parameters of the labor supply model.

The vector Z in (4) may include exogenous variables like age, gender and education, but can also contain past labor market states, and measures for the incentive effects of social security, pension plans and health insurance coverage. There is a substantial literature that deals with the importance of financial incentives in explaining retirement decisions (see Lumsdaine and Mitchell, 2000, for an extensive overview). There is a consensus that financial incentives are important determinants for the retirement decision and that generous retirement schemes of the social security and pension system have induced many older workers to retire early (see, for example, Blondal and Scarpetta, 1998; Gruber and Wise, 1998).[4] For the Netherlands there is strong evidence that a substantial part of the massive drop in the participation rate of older workers has been caused by incentives provided in the Dutch institutions. Kapteyn and de Vos (1998) calculated implicit tax rates for ER, UI and DI schemes in the Netherlands. In the early 1990s the maximum implicit tax rates of UI and DI schemes are about 60% and peak at age 58, and implicit tax rates of the ER schemes in the Netherlands sometimes even exceeded 100%. ER benefit levels were on average 80% of last gross earnings and the replacement rates did not increase if enrollment into the programs was postponed. This implies that the implicit tax rates associated with these schemes 'peak' (i.e., are financially most attractive; see Coile and Gruber, 2000) at the very moment that a worker becomes eligible for an ER scheme. This implies that for our empirical application, which uses data for 1991, 1993 and 1995, it will be important to deal explicitly with the individual's eligibility status for an ER scheme (observed in our data).

In the empirical model, the transition rates in (4) are specified as multinomial logit probabilities with random individual effects. If an individual is eligible for an early retirement scheme at time t, the transition probabilities are defined as

$$\begin{aligned}&\Pr(S_{it} = j | S_{i,t-1} = \text{Work}, \mathrm{E}(D_{it}^{\text{Work}}), Z_{it}, \xi_i^{\text{UI}}, \xi_i^{\text{DI}}, \xi_i^{\text{ER}})\\ &\quad = \frac{\exp(\gamma^j \cdot \mathrm{E}(D_{it}^{\text{Work}}) + Z_{it}\delta^j + \xi_i^s)}{\sum\limits_{l\in\{\text{Work},\text{UI},\text{DI},\text{ER}\}} \exp(\gamma^l \cdot \mathrm{E}(D_{it}^{\text{Work}}) + Z_{it}\delta^l + \xi_i^l)}, \quad \text{for } j = \text{Work, UI, DI or ER} \quad (4')\end{aligned}$$

If the respondent is not eligible for early an retirement scheme, the transition probabilities are

$$\begin{aligned}&\Pr(S_{it} = j | S_{it-1} = \text{Work}, \mathrm{E}(D_{it}^{\text{Work}}), Z_{it}, \xi_i^{\text{UI}}, \xi_i^{\text{DI}})\\ &\quad = \frac{\exp(\gamma^j \cdot \mathrm{E}(D_{it}^{\text{Work}}) + Z_{it}\delta^j + \xi_i^s)}{\sum\limits_{l\in\{\text{Work},\text{UI},\text{DI}\}} \exp(\gamma^l \cdot \mathrm{E}(D_{it}^{\text{Work}}) + Z_{it}\delta^l + \xi_i^l)}, \quad \text{for } j = \text{Work, UI or DI} \quad (4'')\end{aligned}$$

[4] Employer-provided private insurance is not relevant in the Netherlands. We will therefore not deal with this issue in our empirical model.

DOI: 10.1002/jae

The reference category is Work, so the normalization restrictions are: $\gamma^{\text{Work}} = \delta^{\text{Work}} = 0$ and $\xi_i^{\text{Work}} = 0$, for all i. By including the eligibility status in the transition probabilities, we control for the change in the financial incentives once an ER scheme becomes available for an individual. In addition, we will include wages and benefit levels (to be discussed in more detail in Section 4) in the set of controls (Z).

2.4. Health Production

The labor supply model ($4'$) and ($4''$) can now be jointly estimated with equation ($1'$). This still leaves the effect of work on health unspecified. To deal with the endogeneity of H^o, we extend the labor supply model and the health reporting model with a model for the impact of work (history) on observed health. For this 'health production' model (Grossman, 1972) we use a random effect linear regression specification:

$$\text{pdf}(H_{it}^o | \{S_{i\tau}\}_{\tau=0}^{t-1}, X_{it}^o, \eta_i^o) = \varphi(\frac{f(\{S_{i\tau}\}_{\tau=0}^{t-1}) + X_{it}^o\theta + \eta_i^o}{\sigma}) \tag{5}$$

where φ is the probability density of the standard normal distribution. The function f is a linear function of indicators summarizing the respondent's labor market history. We will be more specific about the exact specification in Section 4.

The full model for the interrelation between health and work now consists of the labor supply equations ($4'$) and ($4''$), the health reporting equation ($1'$) and the health production equation (5). These equations will be estimated jointly. Joint estimation is required here because the three equations may be stochastically related by means of their unobserved components. Moreover, there exist parameter restrictions across equations because the cleansed health measure E(D^{work}) used in the labor supply equations is derived from the reporting equation.

The model contains five time-invariant unobserved individual-specific effects: ξ_i^{UI}, ξ_i^{DI} and ξ_i^{ER} for the labor market states, $\eta_i{}^*$ for the reporting model and η_i for the health production model. We will estimate this model by the simulated maximum likelihood method. The Appendix provides details regarding the construction and estimation of the likelihood function.

2.5. Identification

The complete model describes how health affects labor supply decisions (equations ($4'$) and ($4''$)). In modeling this causal effect of health on labor supply, the potential endogeneity of the health variables is taken into account by explicitly modeling two of its sources: the reverse causal effect of labor market status on health and the potential state dependence of subjective health measures. The identification of these mechanisms is achieved by two assumptions. Firstly, it is assumed that the effect of the labor market status on health is a gradual process rather than an instantaneous effect. This can be seen in equation (5), in which the individual's work history is included among the explanatory variables, but its current labor market status is not. Secondly, the mechanism describing the state dependence of the subjective health measure is identified by Assumption 1, discussed above.

Furthermore, in the health reporting equation and the multinomial logit model for labor market transitions we need some arbitrary normalization to identify the parameters. In the health reporting equation (2) there is no intercept in the index function and the variance of e^* is set equal to one; in the transition probabilities (4) the labor market state Work is used as the reference category.

Of prime interest in the reporting model is the relation between labor market status and health. Exogenous variables might also affect reporting behavior directly; i.e., the reporting thresholds in (1) may also depend on exogenous characteristics ($c_k = c_k\ (S, X)$). In Kerkhofs and Lindeboom (1995) the hypothesis that the exogenous variables affect the thresholds differently is tested and rejected. This, however, still leaves the possibility that an exogenous variable shifts all the thresholds by the same amount, and a shift like that cannot be distinguished from an opposite shift of the index function in (2). So, if an exogenous variable in X has a positive effect on health (i.e., shifts the index down), this is observationally equivalent to an equal-sized shift of the thresholds to the right, which is tantamount to a more positive way of reporting on the ability to work. For instance, if respondents with a higher education level report fewer limitations due to health problems, this may mean that they are in fact healthier, but it may also mean that they are not really healthier, but tend to be more positive in reporting about their health status. Without further assumptions or more information, this identification problem cannot be solved. As our purpose was only to identify differential reporting with respect to the endogenous variable S, this is not essential, but it means that we have to make an additional assumption to make the model identified. Here we have chosen to assume that the variables in X only affect the index in (2) and not the thresholds. It has to be clear, however, that this is an identification assumption and that the estimated parameter can signify a real effect on health, but also an effect on the way health is reported.

2.6. Relation to Other Studies in this Area

Bound *et al*. (1999) study the effect of (lagged) health on labor force participation behavior. They address the potential bias in a subjective health measure (H) and account for this by relating this measure to a range of objective health indicators (H^o) and socio-demographic variables. This equation is used to instrument the error-ridden self-reported health measure in the labor supply model. This approach was also followed in Bound (1991). This study is one of the most influential articles in this field. In his model, reporting depends on the financial rewards of retirement.[5] This model is used to compare several of the approaches that were used in the existing literature. Stern (1989) constructs a latent variable model for labor supply, true health and reported health. Labor supply is influenced by unobserved true health, reported health depends on true health and labor market status, and unobserved true health depends on labor force participation. Effectively, he estimates a model for labor force participation, where health is instrumented in the same way as in Bound (1991). Au *et al*. (2004) use the same approach as Bound *et al*. (1999), but in their model labor market status, rather than income, may affect reporting behavior. In their labor supply model they use a similar instrument as in Bound *et al*.; i.e., they construct a model for self-reports as a function of demographic characteristics and the prevalence of chronic conditions and use the estimates of this model to instrument the error-ridden self-reports. Our approach differs in three important ways from these contributions. First, we model state-dependent reporting biases by allowing the labor market status to have a differential effect on the threshold levels in (2). This

[5] We focus on reporting differences that depend on labor market status. This is appropriate in the context of the Dutch disability income arrangements in the 1990s. Dutch disability rates were very high and there is evidence (Aarts and de Jong, 1992) that DI schemes were used for early retirement. Moreover, Dutch DI, UI and ER benefits are a fixed fraction of the last wage and therefore there is little variation in this variable.

DOI: 10.1002/jae

allows for a more flexible effect of the labor market status on the distribution of the categorical limitations variable H. Secondly, we make the assumptions underlying the identification of the reporting bias explicit and we identify the extent of the bias for non-working individuals. Thirdly, in our model, 'objective' health is modeled as an endogenous variable and the health production model yields estimates of the determinants of health, including the effect of work on health.

Our way of dealing with justification bias is similar to Kerkhofs and Lindeboom (1995) and Kreider (1999). Kerkhofs and Lindeboom (1995) use a partial model that takes labor market status and health indicators as exogenous. Kreider (1999) estimates his reporting model jointly with a model for work status, but assumes exogeneity of the health indicators. Furthermore, he implicitly makes an assumption equivalent to Assumption 1, in order to account for reporting biases. Kreider and Pepper (2008) develop a non-parametric bounding framework for assessing how assumptions on the reporting process affect inferences about the health effects on employment.

Benitez-Silva *et al.* (2004) test the hypothesis of rational unbiased reporting of disability status. Their test is based on a comparison between award decisions by the disability insurance administrators (SSA) and a self-reported disability measure. If individuals exaggerate their health problems, one would expect the rate of self-reported disability to exceed the fraction of those who are ultimately awarded benefits. They test the hypothesis on US data and cannot reject the hypothesis that self-reported disability is an unbiased indicator of the SSA's award decision.

Some studies do not address the issue of reporting bias in self-assessed health, but analyze work and health simultaneously. Sickles and Taubman (1986) develop a joint model for health and retirement, in which health is instrumented with a set of exogenous variables. Sickles and Yazbeck (1998) specify and structurally estimate a model of the demand for leisure and the production of health. Simultaneity is taken into account by explicitly modeling the effect of health on work and the effect of work (history) on health, allowing for possibly correlated unobserved individual effects (the ηs and ξs in our model).

3. DATA

The data for the empirical application are obtained from the first two waves of the Leiden University Center for Research on Retirement and Aging (CERRA) panel survey. The CERRA panel survey is a Dutch survey, specifically designed for the analysis of the labor market behavior of the elderly. It resembles the American Health and Retirement Study (HRS). The first wave was fielded in the fall of 1993 and consists of 4727 households in which the head of the household (defined as the main income earner) was between 43 and 63 years of age at the date of the interview. In each household both head and partner, if present, were interviewed. The respondents also had to complete a written questionnaire. In the fall of 1995 the same respondents were contacted for a second interview. Approximately 74% of the first wave respondents participated in this second wave, which resulted in about 3500 households. For each wave extensive information is obtained about work history and current labor market status, sources of income, attitude towards retirement, ER eligibility, health and a variety of socio-economic variables.

In the subsequent analyses we focus on the male heads of households, excluding the self-employed. After removal of inconsistent observations and observations with missing values on the key variables, we obtain a dataset of 3038 individuals.

The *health variables* in the CERRA data include the commonly used subjective measures based on responses to questions like 'How good would you rate your health?' and 'Does your health

DOI: 10.1002/jae

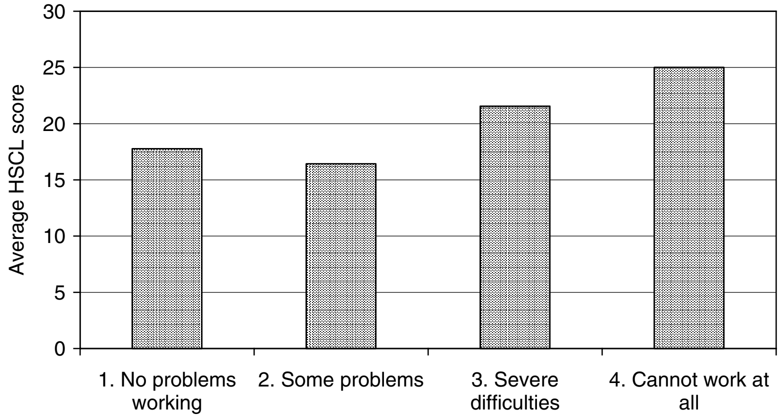

Figure 1. Average HSCL score relative to work-related health condition

limit you in your ability to work?' Responses to the latter question are used as the subjective work-related health measure H in this paper. The written questionnaire contains the complete set of 57 items of the Hopkins Symptoms Checklist (HSCL).[6] The responses to these questions result in a total health score, which can take on an integer value between 0 (best health state) and 171 (worst health state). As motivated in Section 2.2, we use the total HSCL score as the general health measure H^o in the empirical analysis. The advantage of this measure is that, by its construction, it is less sensitive to justification bias than responses to the work limitations question. The HSCL score contains a large number of different items, does not refer to work or benefits, and it was conducted in the written part of the survey (no interviewer present).

Figure 1 shows the average HSCL score for each of the response categories of the work-related health variable H. For both health indicators, higher values indicate a poorer health. The figure shows a clear positive relation between both health measures; respondents who claim that health limits their ability to work have on average higher HSCL scores. For the respondents whose health causes only some problems for paid work, the HSCL score is slightly below that of respondents who experience no problems at all, although this difference is small and not statistically significant.

Table I shows how both health variables are related to labor market status. Respondents who are in the disability scheme on average have considerably higher scores for both health variables. The unemployed have a slightly higher HSCL score than the employed, but their probability of

Table I. Average HSCL score (H^o) and subjective work-related health score (H) in 1993

Labor market state in 1993	Average HSCL score (H^o)	Work-related health indicator (H), (% with severe difficulties or who cannot work at all)
Workers	10.0	5.6
Unemployed (UI)	11.9	23.9
Disabled (DI)	24.8	92.2
Early retired (ER)	8.0	12.9
All	12.3	23.3

[6] See Derogatis *et al.* (1974) for more information on the items of the test score and validation results.

DOI: 10.1002/jae

experiencing severe difficulties or not being able to work at all is four times as large as that for the employed. Interestingly, the HSCL score shows that on average the health condition of the early retired is better than for workers, although they report more health-related problems with respect to their ability to perform work. Table A.I contains some further descriptive statistics of the dataset that is used in the empirical application.

4. EMPIRICAL ANALYSES

The equations for work ($4'/4''$), health reporting ($1'$) and health production (5) are related by means of the correlated random effects and cross-equation restrictions on the parameters (the cleansed health measure $E(D^{\text{Work}})$, defined in (3), enters the model for work ($4'/4''$)). The random effects are drawn from a five-dimensional normal distribution. The three sub-models are estimated simultaneously by maximizing the simulated likelihood.

4.1. Labor Force Participation

The model for work decisions is a mixed multinomial logit, with the employed category as a reference group. Table II reports the results. A positive coefficient is associated with a higher

Table II. Estimates of the full model for work, health and health reporting. Part I: labor supply and retirement[a]

	Labor market states[b]					
	Unemployed		Disabled		Early retirement scheme	
	Parameter	*t*-value	Parameter	*t*-value	Parameter	*t*-value
Constant	−10.561	−2.70	−11.745	−2.58	2.441	1.56
Age = 56	1.033	2.27	0.144	0.21	0.580	1.48
Age = 57	0.700	1.34	−0.304	−0.40	1.091	2.78
Age = 58	0.270	0.54	−1.006	−1.41	0.520	1.38
Age = 59	1.002	1.93	0.051	0.07	1.681	4.29
Age = 60	0.339	0.59	−0.425	−0.54	1.280	3.21
Age = 61	0.879	1.43	0.158	0.19	2.056	4.34
Age 62 and older	1.642	2.82	−0.895	−1.21	1.252	2.64
Education medium	1.307	4.13	0.749	1.65	0.667	2.61
Education higher	1.270	3.66	0.396	0.83	0.469	1.79
Mean annual benefit income[c]	7.699	1.77	8.154	1.66	−1.526	−0.98
Health index[d]	21.740	8.51	29.367	9.76	14.649	6.28
Transition variables (from working at $t-1$)						
Constant	−3.607	−5.77	−1.729	−1.03	−2.721	−5.37
Part-time work (<32 h p.w.)	−0.365	−0.68	1.016	1.27	−1.442	−3.74
Route-specific replacement rate[e]	0.493	0.76	5.120	1.27	0.086	0.62
ER scheme available in firm	−0.107	−0.45	−0.748	−2.61		
Log-likelihood (complete model)	−11157.48					

[a] The model is estimated with the simulated maximum likelihood method (50 draws).
[b] The reference category is paid work.
[c] Mean annual benefit income (route-specific) until the mandatory retirement age (logarithm).
[d] Probability of reporting bad health corrected for state-dependent reporting differences (i.e., $E(D^{\text{Work}})$ from equation (3)).
[e] Total benefit incomes up to the age of mandatory retirement divided by total wage income associated with continued work (logarithm).

probability of being in a particular labor market state. The upper panel of the table refers to control variables at time t; these include a flexible function of age (with separate dummy variables for ages of 56 and higher), education, mean annual retirement income and the cleansed health measure $E(D^{\text{Work}})$. The lower panel of the table includes a set of financial variables that are expected to be of direct relevance for the transition out of work. One such variable is the logarithm of the replacement rate, which measures the total value of the benefits for a specific exit route relative to the total value of wage income associated with working until the mandatory retirement age of 65. For an ER scheme this variable measures the effect of the level of the benefits *conditional* on eligibility for the ER scheme.[7] The benefit income for each exit route is based on estimates of an auxiliary panel data model for wage dynamics (see Heyma (2004) for details) and specific rules of the social security and early retirement schemes (see Kapteyn and de Vos, 1998, for a description of Dutch social security and ER schemes).

The coefficients of the financial incentive variables have the right sign, but are not significant for any of the exit routes. For UI and DI, this may be due to the relatively small number of transitions that we observe through these channels. Moreover, there is relatively little variation in the replacement rates as a substantial fraction of new UI and DI recipients qualify for wage-related benefits of 70% up to the statutory retirement age. We have also tried alternative specifications of the incentives for entrance into a UI and DI scheme, such as the maximum income gain from postponing retirement as opposed to immediate retirement (the option value; Stock and Wise, 1990). These specifications did not improve the results. Apparently, much of the reward of postponing retirement is captured by the dummy variable indicating the availability of an ER scheme and the flexible age specification.

The small and insignificant effect of the financial incentive variable for the ER scheme is due to the fact that this variable measures the incentive effect, *conditional* on eligibility for the ER scheme. As mentioned above, the replacement rates of the Dutch ER schemes that prevailed in the early 1990s were very high and the implicit tax rate 'peaked' at the very moment that an individual would become eligible for an ER scheme. This induced a strong incentive to retire immediately upon eligibility.[8]

In the lower panel of the table, for the transitions to UI and DI, we have included an indicator for the availability of an early retirement scheme within the firm. This variable is included to capture the effects of the availability of a very generous future retirement option. The parameter estimate for this variable is negative and significant for the DI exit route, indicating that the availability of an ER scheme lowers the probability that an individual enters into a DI scheme. Hence, DI schemes and ER schemes appear to act as competing retirement options. No such effects are found for UI schemes. It is important to note that both ER and UI schemes are financed from sector funds and that it was common practice for employers to provide laid-off workers a bonus on top of their UI benefit up to the time that they qualified for an ER scheme. This means that the two schemes were sometimes used in conjunction, acting as complements rather than substitutes.[9]

Of particular interest for this paper is the effect of health on work. The cleansed health index $E(D^{\text{work}})$ is used to instrument the error-ridden self-report H. For each exit route, the estimates in

[7] The eligibility status of the individual influences the denominator of the multinomial logit ($4'/4''$).

[8] In the data one can see that about 80% of those who became eligible for an ER scheme retired immediately and the remaining 20% retired within 1.3 years. Calculation with the model show that the probability of entering into an ER scheme is relatively high (cf. the constants in Table II).

[9] The sector funds are formed from employer and worker contributions and until recently there was no experience rating. This implies that it will in general be difficult to distinguish between the separate incentive effects of ER and UI schemes.

DOI: 10.1002/jae

Table II show strong and significant effects of health on labor market behavior. All coefficients are positive, implying that poor health is associated with higher probabilities of not working. To assess the impact of justification bias, we should know what these estimates would be if we would not correct for state-dependent reporting and would directly include the dummy variable D. The estimated health effects in this 'simple' model are 0.826, 4.179 and 0.511 for UI, DI and ER, respectively (full table is available upon request from the authors). These estimates, as expected, indicate that health is by far the strongest factor for people on a DI scheme. In order to compare these estimates to those in Table II, we have rescaled the health coefficients in Table II by the average difference in the value of the cleansed health measure between respondents reporting bad health and respondents reporting good health.[10] After rescaling, the health coefficients of Table II become 2.131, 2.261 and 0.571, for UI, DI and ER, respectively. Comparison of these numbers with the numbers from the 'simple' model reveals that for DI recipients the use of the biased self-reports on health leads to exaggerated effects of health on work (the coefficient drops from 4.179 to 2.261 if justification bias is accounted for). For UI recipients the effect of health becomes more pronounced, whereas it remains approximately the same for ER recipients. We have to turn to the results of the reporting model to better understand these results.

4.2. Health Reporting

The response to the survey question about health-related work limitations is a categorical variable with values ranging from 1 (no problems at all) to 4 (cannot work at all). We use a mixed ordered probit model to explain the responses. The index function, representing latent health, includes a third-order polynomial in the HSCL score (H^o) and additional controls such as age, white-collar work, education and quadratic functions of the total number of months worked up until t and the number of months worked in the last 10 years. The thresholds of the ordered response model depend on labor market status S and, if assumption 1 holds, the effect of S on the thresholds can be interpreted as reporting bias. For reasons discussed in Section 2.5, we did not include exogenous variables in the thresholds and cannot therefore distinguish the causal effects of these variables on health from the way health is reported. We therefore have to be cautious with the interpretation of the effects of these variables in the index function.

Table III reports the results. The first threshold ${\psi_1}^{\mathrm{Work}}$ for workers is estimated at -0.568. The first threshold for the other labor market states are relative to this value (${\psi_1}^k - {\psi_1}^{\mathrm{Work}}$, $k =$ UI, DI, ER). The other threshold parameters are presented as increments: ${\psi_{j+1}}^k - \psi_j^k$, $k =$ Work, UI, DI, ER, $j = 1, 2$. The threshold values can be computed from these parameters and are presented graphically in Figure 2.

The threshold values demonstrate that there are strong reporting effects for DI recipients. Conditional on the value of the latent health index, DI recipients significantly more often report that health problems reduce their ability to work than respondents in the other three groups. A latent health index above the second threshold level signifies that respondent will report that their health prevents them working or causes severe difficulties doing so. For workers, unemployed and early retired the values of this second threshold are close together. For the DI recipients it is much lower.

[10] The weights that are used to scale the parameter estimates of health in Table II are 0.098, 0.077 and 0.039, for UI, DI and ER, respectively.

DOI: 10.1002/jae

Table III. Estimates of the full model for work, health and health reporting. Part II: health reporting[a,b]

	Subjective work-related health	
	Parameter	t-value
(i) Control variables		
Age = 56	0.045	0.55
Age = 57	0.201	2.49
Age = 58	0.186	2.29
Age = 59	0.104	1.22
Age = 60	0.156	1.76
Age = 61	0.040	0.46
Age 62 and older	−0.028	−0.35
Partner	0.040	0.88
White-collar worker	−0.125	−3.74
Education medium general	−0.218	−3.73
Education medium vocational	−0.173	−3.27
Education higher general	−0.048	−0.60
Education higher vocational	−0.198	−3.53
Education academic	−0.191	−2.24
Number of months worked in past 10 yr[c]	0.396	0.99
Number of months worked in past 10 yr squared[c]	−1.857	−6.48
Number of months ever worked[c]	−0.204	−2.33
Number of months ever worked squared[c]	0.044	3.36
(ii) General health measure (H^o)[c]		
Total score HSCL	1.718	3.82
Total score HSCL squared	−5.357	−3.54
Total score HSCL cubed	3.109	2.36
(iii) Reporting behavior threshold levels		
Work: $\psi_1{}^{Work}$	−0.568	−4.67
$\psi_2{}^{Work} - \psi_1{}^{Work}$	0.657	18.25
$\psi_3{}^{Work} - \psi_2{}^{Work}$	0.754	11.51
UI: $\psi_1{}^{UI} - \psi_1{}^{Work}$	0.394	4.05
$\psi_2{}^{UI} - \psi_1{}^{UI}$	0.327	6.77
$\psi_3{}^{UI} - \psi_2{}^{UI}$	0.441	6.84
DI: $\psi_1{}^{DI} - \psi_1{}^{Work}$	−1.282	−10.45
$\psi_2{}^{DI} - \psi_1{}^{DI}$	0.347	5.28
$\psi_3{}^{DI} - \psi_2{}^{DI}$	1.021	15.40
ER: $\psi_1{}^{ER} - \psi_1{}^{Work}$	0.339	4.54
$\psi_2{}^{ER} - \psi_1{}^{ER}$	0.453	11.53
$\psi_3{}^{ER} - \psi_2{}^{ER}$	0.704	11.41
Log-likelihood (complete model)	−11157.48	

[a] Subjective health is a four-category ordinal variable derived from the question 'Does your health limit you in the kind and amount of work that you can do?' with answers ranging from 'No problems' to 'Cannot work at all'.
[b] The model is estimated with the simulated maximum likelihood method (50 draws).
[c] HSCL score is scaled by 125; number of months is scaled by 120.

There are also some differences between UI recipients and early retired. All thresholds of the early retired are slightly above those of the employed. This implies that the early retired will be a little more cheerful about what health does and does not allow them to do than the workers. Compared to these two groups, the threshold levels for the UI recipients are less dispersed, implying that for the same state of health the response distribution of the unemployed will have

DOI: 10.1002/jae

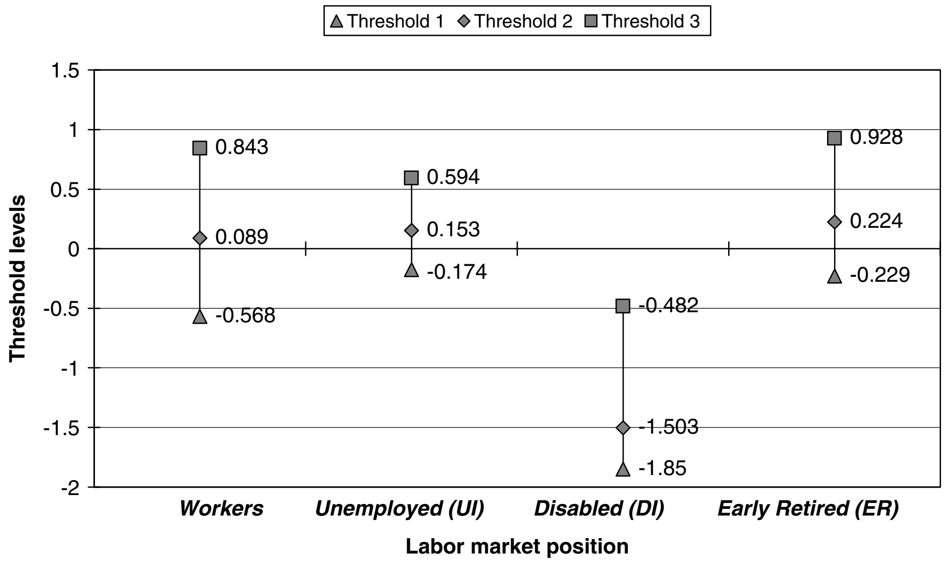

Figure 2. State dependence of thresholds in ordered probit health reporting equation. If the index of ordered probit model lies: below threshold 1—no problem working; between thresholds 1 and 2—some problems; between thresholds 2 and 3—severe difficulties; above threshold 3—cannot work at all

heavier tails, with a larger proportion reporting they cannot work at all, but also a larger proportion reporting no problems whatsoever. These findings are in line with the findings of the previous subsection. There we concluded that after correcting for state-dependent reporting behavior we observe smaller health effects for DI recipients, larger effects for UI recipients and approximately the same effects for ER recipients.

To gain some further insight into these results, we have used the parameter estimates to predict whether respondents would state that health restricts their ability to work and—for the UI, DI and ER groups—to predict what their response would have been if they would have been employed, keeping everything else the same. So, for each respondent we can now compare their actual answer to a predicted response after correcting for justification bias (i.e., the 'cleansed' self-report $E(D^{\text{Work}})$). The results of these calculations are shown in Figure 3.

A comparison of the actual (observed) and predicted frequency distribution of H indicates that the model describes the data fairly well. A comparison of the actual responses with the cleansed health measure confirms what we already saw in Figure 2: there are no major changes for ER recipients and the changes for UI recipients are small. For the DI recipients the picture is quite different: the reporting biases appear to be large and systematic. DI recipients respond much less favorably about the state of their health, but the frequency distribution is almost reversed if one corrects for state-dependent reporting. More specifically, about 60% of the disabled workers report that they cannot work at all, but after correcting for the reporting bias this proportion is reduced to about 20%. This result is less extreme than it appears and may even be in line with other findings. Burkhauser *et al*. (1999) note that work accounts for the larger part of the income of men aged 51–61 in the USA, while for the Netherlands a large percentage of the men in this age category have already withdrawn from the labor market. They argue that differences between Dutch and US institutions, rather than health differences, can explain this difference. Furthermore, about 24%

DOI: 10.1002/jae

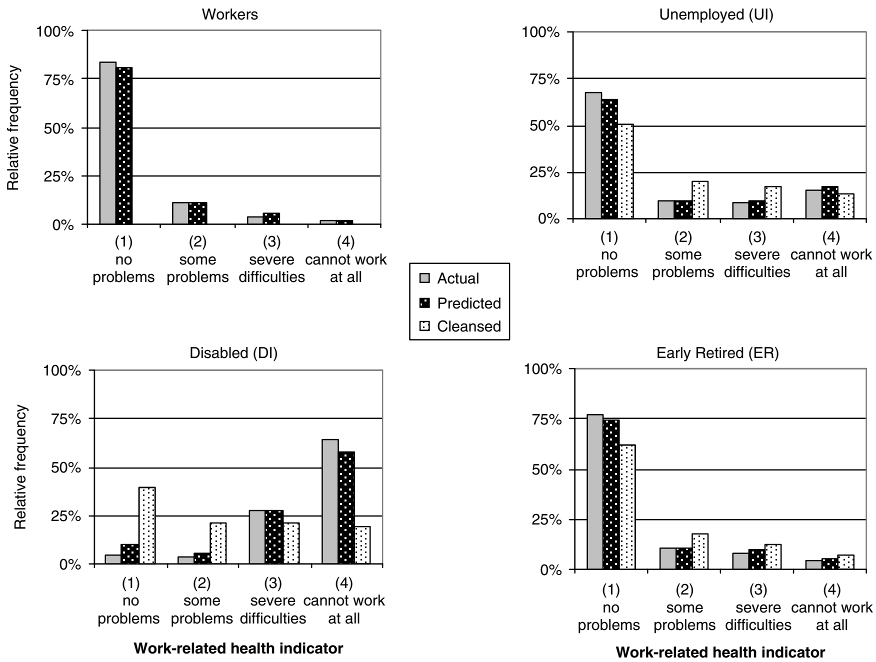

Figure 3. Sample distribution of observed and predicted values of the work-related health measure (H) and 'cleansed' predictions accounting for justification bias

of the older (above 55) male Dutch workers are in a DI scheme. This is more than three times the corresponding DI rate in the USA (7%).

One could argue that the HSCL score may to some extent also suffer from justification bias. If this were the case, the estimated reporting effect for DI recipients would be an under bound for the total effect of justification bias. We have also estimated a simple ordered probit model, in which we did not take the endogeneity of the labor market status and the HSCL score into account (η^* excluded). The estimate for the lowest threshold ($\psi_1^{\mathrm{DI}} - \psi_1^{\mathrm{Work}}$) now equals -1.923 (t-value $= 16.70$); it was -1.282 in Table III.[11] Thus these estimates suggest an even larger share of healthy DI recipients.

Concerning the effects of the variables in the index function of the ordered response specification, it can be seen that not only the general health indicator (HSCL score, H^o), but also personal characteristics and work history, are important determinants. As noted in Section 2.5, these coefficients have to be interpreted with caution, as they may explain differences in health, but also differences in health reporting.

4.3. Health Production: The Impact of Work on Health

Table IV reports the results for the health production model. Low values of the HSCL score correspond to good health and high values to bad health. The older age groups appear to be

[11] The second and third incremental values for the DI in this model are $\psi_2^{\mathrm{DI}} - \psi_1^{\mathrm{DI}} = 0.324$ ($t = 5.66$) and $\psi_3^{\mathrm{DI}} - \psi_2^{\mathrm{DI}} = 0.970$ ($t = 16.36$). The full table is available upon request from the authors.

DOI: 10.1002/jae

Table IV. Estimates of the full model for work, health and health reporting. Part III: health stock equation[a]

	HSCL score[b]	
	Parameter	t-value
Constant	3.085	33.38
Age = 56	−0.023	−0.29
Age = 57	−0.162	−2.14
Age = 58	−0.118	−1.45
Age = 59	−0.348	−4.18
Age = 60	−0.282	−3.50
Age = 61	−0.465	−5.29
Age 62 and older	−0.491	−6.64
Partner	−0.467	−7.82
Family size	−0.015	−0.67
Education medium general	−0.052	−0.76
Education medium vocational	−0.119	−2.08
Education higher general	−0.036	−0.37
Education higher vocational	−0.023	−0.39
Education academic	0.143	1.49
Mean annual income[c] (logarithm)	−0.105	−1.82
At work in $t-1$	−0.052	−0.49
Number of months ever worked[d]	−0.468	−3.36
Number of months ever worked squared[d]	0.090	4.24
Number of months worked in past 10 yr[d]	−0.670	−0.68
Number of months worked in past 10 yr squared[d]	0.086	0.05
Number of months worked in past 10 yr cubed[d]	−0.363	−0.34
σ^2 (variance)	1.293	56.23
Log-likelihood (complete model)	−11157.48	

[a] The full model is estimated with the simulated maximum likelihood method (50 draws).
[b] High values are associated with bad health.
[c] Mean of annual income associated with working until the mandatory retirement age.
[d] Number of months scaled by 120.

healthier than the reference group (younger than 55). This somewhat unexpected pattern also shows up in the raw data and is not caused by the model specification. It is good to note that these differences are quantitatively small (the average HSCL score for the whole sample is 12.34). Moreover, the variable captures differences between age groups as well as differences between birth cohorts. The majority of the younger respondents (the reference category) are born shortly before or during the World War II. By now there is an extensive literature on the effects of early childhood conditions on later life health and mortality (see, for instance, the recent surveys in Doblhammer, 2004, or Case *et al*., 2005). This literature concludes that exposure to malnutrition and/or infectious diseases during early childhood may increase the probability of ill health in later life. Lumey and Van Poppel (1994) and Roseboom *et al*. (2001) find that the health condition of individuals who grew up in the World War II and in particular during the 1944 famine at the end of World War II have a poorer state of health at later ages. In Kerkhofs and Lindeboom (1997) we have analyzed the HSCL test scores using fixed effects regressions and also found that cohorts born shortly before or during World War II had less favorable health than the earlier and later birth cohorts.

The results of the other variables are in line with earlier findings in economics and epidemiology. Respondents with a partner are healthier than those without a partner. The education variables suggest that the reference group (those with only primary school) is the least healthy. We also find that higher incomes are associated with better health. The effect of work history (the function $f(.)$ in equation (5)) may be particularly interesting from a policy point of view, as this may be informative on the consequences of prolonged work efforts for the average health conditions of future retirees. We have captured the individual's work history by a third-order polynomial in the total number of months worked up until t and the number of months worked in the last 10 years. The latter is included to capture effects from recent work efforts on health.[12] To illustrate the quantitative effect of work history on health, we computed the HSCL profile of an average worker with about 15 years of work experience who from then on continues to work for a number of years, controlling for the effect of aging and (changes in) other characteristics. The computation shows that after about 300 months (25 years) of work general health starts to deteriorate with work (higher HSCL scores). The effect is quantitatively important: the HSCL score increases from about 8 (300 months) to about 14 (600 months). The sample mean of the number of months worked about is 350 months and more than 75% of the sample has worked over 300 months, so that effectively for most individuals continuing to work is at some stage associated with worse health outcomes.

5. CONCLUSIONS

This paper explores the relationship between health and work labor supply of workers. For such an analysis two issues are relevant. Health and work are simultaneous variables in the sense that there are direct effects of health on work and vice versa, and because there are common unobservables that affect observed health and work outcomes. Furthermore, self-assessed health measures are often used for this type of empirical analyses, but some research indicates that these may be affected by endogenous, state-dependent reporting behavior. The novelty of this study is that we specify and estimate a simultaneous model that takes all these issues into account. The model consists of three parts that are linked by correlated random effects. The first part of the model describes labor supply decisions, in particular how financial incentives and health affect retirement behavior. The second part models the state of health and how this is related to work life. The third part of the model describes how the labor market status affects observed self-reports on health, particularly on how it limits the ability to work. This third part of the model is used to assess the impact of endogenous misreporting and to compute an adjusted work-related health measure, which is stripped from state-dependent misreporting. This cleansed health measure is used in the labor supply model. The three models are estimated jointly on Dutch data using simulated maximum likelihood techniques.

We find that financial incentives are important for the decision to stop working. Workers in the Netherlands have strong incentives to take an employer-provided early retirement (ER) scheme as soon as they become eligible for this scheme. The eligibility for an ER scheme substantially reduces the probability of early outflow through a disability insurance (DI) scheme. Hence, it appears that ER and DI schemes act as substitutes. Furthermore, we find strong effects of health on work choices. The estimated health production equation reveals that at older ages increased work efforts lead to a deterioration of health. This suggests that pension and social security reforms

[12] We do not observe how many hours people have worked in the past and therefore the work history variables reflect the extensive margin rather than the intensive margin.

DOI: 10.1002/jae

that aim at increasing labor force participation rates of older workers may have an adverse effect on the distribution of health among the elderly, with obvious consequences for pension and health care policies.

Finally, the results show that justification bias is substantial and that failing to account for this may change estimation results considerably.

6. APPENDIX A

Table A.I. Sample descriptives

	Mean	Min.	Max.
Age in 1993	55.13	42.25	65.00
Education medium general	0.126	0	1
Education medium vocational	0.163	0	1
Education higher general	0.046	0	1
Education higher vocational	0.179	0	1
Education academic	0.059	0	1
White-collar worker	0.538	0	1
Family size	2.646	1	10
Number of months ever worked	347	0	707
Early retirement scheme in firm	0.716	0	1
Unemployed (UI)	0.073	0	1
Disabled (DI)	0.176	0	1
On ER scheme	0.195	0	1
Reports bad health[a]	0.140	0	1
Total HSCL score	12.339	0	125
Attrition between wave I and II	0.24		
Number of observations in wave I	3038		

[a] Whether the respondent reports to have difficulties with performing work or that he/she cannot work at all.

Table A.II. Estimates of the full model for work, health and health reporting. Part IV: labor force participation in 1991[a,b]

	Parameter	t-value
Constant	1.322	2.72
Age = 56	−0.825	−4.58
Age = 57	−1.366	−8.06
Age = 58	−1.436	−8.25
Age = 59	−2.293	−11.17
Age 60 and older	−3.047	−12.54
Education medium	0.764	6.09
Education higher	0.848	5.76
Income of continued work	−0.020	−0.13
Eligible for ER scheme	−1.082	−7.88
Log-likelihood (complete model)	−11157.48	

[a] Logit model with random individual effect for the probability of being at work in 1991. Negative coefficients are associated with lower probabilities.
[b] The full model is estimated with the simulated maximum likelihood method (50 draws).

Table A.III. Estimates of the full model for work, health and health reporting. Part V: parameters of the mixing distribution[a]

	Parameter	t-value
λ	1.620	14.90
u_{11}	2.428	9.16
u_{12}	0.316	1.06
u_{13}	0.708	3.02
u_{14}	0.361	12.95
u_{15}	−0.592	−5.58
u_{22}	−2.100	−8.28
u_{23}	−0.773	−4.16
u_{24}	−0.344	−13.08
u_{25}	−0.032	−0.30
u_{33}	0.157	0.83
u_{34}	0.202	10.07
u_{35}	0.012	0.10
u_{44}	−0.104	−5.50
u_{45}	0.125	0.88
u_{55}	0.225	1.56
Log-likelihood (complete model)	−11157.48	

[a] The full model is estimated with the simulated maximum likelihood method (50 draws).
[b] λ is the factor of proportionality relating the individual effect in the health stock equation (η^o) to the random effect in the health reporting equation (η^*), i.e., $\eta_i^o = \lambda \cdot \eta_i^*$. The parameters u_{11} to u_{55} are the elements of the upper triangular matrix of the Choleski decomposition of the covariance matrix of the random effects. The five remaining random effects (ξ_i^{UI}, ξ_i^{DI}, ξ_i^{ER}, η_i^*, μ_i) are associated with the labor market states UI, DI, ER, the health reporting equation and the initial work state (1991) respectively.

7. APPENDIX B: LIKELIHOOD FUNCTION

Our model describes the health and work states of individuals in 1993 and 1995. The labor supply model includes an indicator for whether an individual was at work in the previous wave. We have an initial conditions problem, since the assignment to the initially observed working state in 1991 may not be random. This could bias the estimates of the labor force participation model. We therefore add a reduced-form equation describing the probability of being at work in the initial year 1991, conditional on exogenous variables V_i and a random individual effect μ_i:

$$\Pr(S_{i,1991} = \text{Work}|V_{i,1991}, \mu_i) = \frac{\exp(V_{i,1991}\omega + \mu_i)}{1 + \exp(V_{i,1991}\omega + \mu_i)} \tag{B.1}$$

The individual effects μ may be correlated with the other unobserved individual effects of the main model. Thus the full model contains six unobserved individual-specific effects: the conditional multinomial logit model (4′/4″) for the labor market status with unobservables ξ_i^{UI}, ξ_i^{DI} and ξ_i^{ER}, the initial conditions equation (B.1), with unobservable μ_i, the health reporting model (1′) with unobservable η_i^* and the health production (5) with unobservable η_i^o. For the latter two we assume a one-factor error specification. More specifically, we take $\eta_i^o = \lambda \cdot \eta_i^*$. This is primarily done to reduce the dimensionality of the model and to ease estimation.

DOI: 10.1002/jae

Conditional on the unobservables (μ_i, $\eta_i{}^*$, η_i^o, ξ_i^{UI}, ξ_i^{DI}, ξ_i^{ER}) the likelihood function is simply the product of the likelihood contributions for the work model ($4'/4''$), the reporting model ($1'$), the health production model (5) and the logit model describing whether or not the respondent had a paid job in 1991 (B.1).[13] In order to compute the likelihood we have to take the expected value w.r.t. the random effects (μ_i, $\eta_i{}^*$, ξ_i^{UI}, ξ_i^{DI}, ξ_i^{ER}):

$$
\begin{aligned}
L = \prod_i \int_{\mu_i} \cdots \int_{\xi^{ER}} & \Pr(S_{i,1991} | V_{i,1991}, \mu_i) \\
& \times \prod_{\substack{t = 1993,\,1995 \\ S_{i,t-2} = \text{Work}}} \{\Pr(S_{it} | S_{i,t-2} = \text{Work}, \mathrm{E}(D_{it}^{\text{Work}}), Z_{it}, \xi_i^{\mathrm{UI}}, \xi_i^{\mathrm{DI}}, \xi_i^{\mathrm{ER}})\} \\
& \times \prod_{t=1993,1995} \{\Pr(H_{it} | H_{it}^o, S_{it}, X_{it}, \eta_i^*) \times \mathrm{pdf}(H_{it}^o | \{S_{i\tau}\}_{\tau=0}^{t-1}, X_{it}^o, \lambda \cdot \eta_i^*)\} \\
& \mathrm{d}F(\xi_i^{\mathrm{ER}}, \xi_i^{\mathrm{DI}}, \xi_i^{\mathrm{UI}}, \eta_i^*, \mu_i)
\end{aligned} \tag{B.2}
$$

In general, no closed-form solution exists and we therefore use simulated maximum likelihood methods to numerically integrate the unobservables out of the likelihood function and maximize the resulting likelihood (Stern, 1997). The basic idea is to draw random variables from the mixing distribution and use these to compute the sample mean of the likelihood function. It is well known that the matrix Σ can be written as UU', where U is an upper-triangular matrix. In order to simulate the unobservables μ_i, $\eta_i{}^*$, ξ_i^{UI}, ξ_i^{DI}, ξ_i^{ER}, we first generate a $5 \times I$ matrix z of independent draws from the standard normal distribution. The columns of the matrix Uz are independent draws from a five-dimensional normal distribution with expectation zero and covariance matrix Σ. The coefficients of U are estimated jointly with the other parameters of the model and are used to calculate the association between the unobservables (μ_i, $\eta_i{}^*$, η_i^o, ξ_i^{UI}, ξ_i^{DI}, ξ_i^{ER}).

For estimation of the likelihood we have used 50 random draws. We also experimented with the number of draws. It appeared that already after 30 draws the results remained very stable. The estimated parameters of the mixing distribution are reported in Table A.III.

ACKNOWLEDGEMENTS

We acknowledge the valuable comments of Rob Alessie, Simon Burgess, Arie Kapteyn, Donal O'Neil, Owen O'Donnell, France Portrait, John Rust, Joao Santos Silva, Marcos Vera, Rainer Winkelmann and participants at the ESSLE conference in Buch am Ammersee, Germany, the Conference on Social Insurance and Pension Research in Arhus, Denmark, the Applied Micro Economics seminar at University College London, participants of the 2nd European Conference on Health Economics held in Lisbon and three anonymous referees.

[13] Retirement is taken as an absorbing state—once out of work, people remain out of work. This is implied by Dutch institutions and confirmed by the data. So, implicitly, non-working individuals are no longer used in the estimation of the labor supply model. They are, however, still used for the estimation of the parameters of the health production and the health reporting model.

REFERENCES

Aarts LJM, de Jong PR. 1992. *Economic Aspects of Disability Behavior*. North-Holland: Amsterdam.

Au D, Crossley T, Schellhorn M. 2004. The effect of health changes and long-term health on the work activity of older Canadians. *IZA Working Paper No. 1281*.

Benitez-Silva H, Buschinski M, Chan HM, Sheidvasser S, Rust J. 2004. How large is the bias in self-reported disability? *Journal of Applied Econometrics* **19**(6): 649–670.

Blondal S, Scarpetta S. 1998. The retirement decision in OECD countries. *OECD Paper No. 202*.

Bound J. 1991. Self reported versus objective measures of health in retirement models. *Journal of Human Resources* **26**: 107–137.

Bound J, Burkhauser RV. 2000. Economic analysis on transfer programs targeted on people with disabilities. In *Handbook of Labor Economics*, Vol. 3c, Arrow KJ, Intrilligator MD (eds). Elsevier: Amsterdam; 3417–3528.

Bound J, Schoenbaum M, Stinebrickner TR, Waidmann T. 1999. The dynamic effects of health on the labor force transitions of older workers. *Labour Economics* **6**: 179–202.

Burkhauser RV, Dwyer D, Lindeboom M, Theeuwes JJM, Woittiez I. 1999. Health, work, and economic well-being of older workers, aged fifty-one to sixty-one: a cross national comparison using the US HRS and the Netherlands CERRA data sets. In *Wealth, Work and Health: Innovations in Measurement in the Social Sciences*, Smith JP, Willis RJ (eds). Essays in Honor of F. Thomas Juster. University of Michigan Press: Ann Arbor, MI; 233–266.

Case A, Fertig A, Paxson C. 2005. The lasting impact of childhood health and circumstance. *Journal of Health Economics* **24**: 365–389.

Coile C, Gruber J. 2000. Social security incentives for retirement. *NBER Working Paper No. 7651*.

Derogatis LR, Lipman RS, Rickels K, Uhlenhuth EH, Covi L. 1974. The Hopkins Symptom Checklist (HSCL): a self-report symptom inventory. *Behavioral Science* **19**(1): 1–15.

Doblhammer G. 2004. *The Late Legacy of Very Early Life*. Demographic Research Monographs. Max Planck Institute for Demographic Research: Rostock.

Grossman M. 1972. On the concept of health capital and the demand for health. *Journal of Political Economy* **80**(2): 223–255.

Gruber J, Wise D. 1998. Social security and retirement: an international comparison. *American Economic Review* **88**(2): 158–163.

Heyma AOJ. 2004. A structural dynamic analysis of retirement behavior in the Netherlands. *Journal of Applied Econometrics* **19**(6): 739–759.

Kapteyn A, de Vos K. 1998. Social security and retirement in the Netherlands. In *Social Security and Retirement around the World*, Gruber J, Wise DA (eds). NBER: Chicago, IL.

Kerkhofs MJM, Lindeboom M. 1995. Subjective health measures and state dependent reporting errors. *Health Economics* **4**: 221–235.

Kerkhofs MJM, Lindeboom M. 1997. Age related health dynamics and changes in labor market status. *Health Economics* **6**: 407–424.

Kreider B. 1999. Latent work disability and reporting bias. *Journal of Human Resources* **34**(4): 734–769.

Kreider B, Pepper JV. 2008. Inferring disability status from corrupt data. *Journal of Applied Econometrics* **23**: 329–349.

Lumey LH, Van Poppel FWA. 1994. The Dutch famine of 1944–45: mortality and morbidity in past and present generations. *Social History of Medicine* **7**(2): 229–246.

Lumsdaine RL, Mitchell OS. 2000. New developments in the economic analysis of retirement. In *Handbook of Labor Economics*, Vol. 3, Ashenfelter O, Card D (eds). Elsevier: Amsterdam; 3261–3307.

McGarry K. 2004. Health and retirement: do changes in health affect retirement expectations? *Journal of Human Resources* **39**(3): 624–648.

Roseboom TJ, van der Meulen JH, Ravelli AC, Osmond C, Barker DJ, Bleker OP. 2001. Effects of prenatal exposure to the Dutch famine on adult disease in later life: an overview. *Twin Research*. **4**(5): 293–298.

Sickles RC, Taubman P. 1986. An analysis of the health and retirement status of the elderly. *Econometrica* **54**(6): 1339–1356.

Sickles RC, Yazbeck A. 1998. On the dynamics of demand for leisure and the production of health. *Journal of Business and Economic Statistics* **16**(2): 187–197.

Stern S. 1989. Measuring the effect of disability on labor force participation. *Journal of Human Resources* **24**(3): 361–395.
Stern S. 1997. Simulation based estimation. *Journal of Economic Literature* **35**(4): 2006–2039.
Stock J, Wise D. 1990. Pensions: the option value of work and retirement. *Econometrica* **58**(5): 1151–1180.
Terza J. 1985. Ordinal probit: a generalization. *Communications in Statistics* **14**(1): 1–11.