



Health shocks, disability and work[☆]



Maarten Lindeboom^a, Ana Llana-Nozal^b, Bas van der Klaauw^{c,*}

^a VU University Amsterdam & Tinbergen Institute, The Netherlands, & HEB, Norway

^b OECD, France

^c VU University Amsterdam & Tinbergen Institute, The Netherlands

HIGHLIGHT

- Exploits unscheduled hospitalizations as unanticipated health shocks in event-history model.
- Health shocks substantially increase disability risk, but no direct effect on employment.
- Finds sizeable (local average treatment) effect of onset of disability on employment.
- Effects larger for men and low-educated workers.

ARTICLE INFO

Article history:

Received 28 August 2015

Received in revised form 20 May 2016

Accepted 23 June 2016

Available online 3 July 2016

ABSTRACT

This paper focuses on the relation between health shocks and the onset of a disability and employment outcomes. We estimate an event-history model using data from the British National Child Development Study (NCDS), where accidents causing a unscheduled hospitalization are the measure for unanticipated health shocks. Our results show that experiencing such a health shock substantially increases the likelihood of the onset of a disability, while it does not have direct effects on employment at later ages. This finding is used to simulate the causal effects of the onset of a disability on later employment outcomes. These simulations show that about two-third of the association between disability and employment can be explained by the causal effect of the onset of a disability on employment. The remaining one-third is selection. For men and lower-educated workers the association is mainly explained by the causal effect, while for women selection is more important.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

A substantial share of the working age population in the industrialized world suffers from a long standing illness or disability that restricts in daily activities and/or work (Dupre and Karjalainen, 2002). Disability prevalence rates are already high at relatively young ages. For instance, in the UK around 5% of the 20–24 year old have a long standing disability and this number increases to around 13% for the 40–44 year old and 28% for those aged 55–59 (Berthoult, 2006). Similar disability rates are found for the US (Kapteyn et al., 2007). Disability is associated with higher benefit take up, poverty and lower employment rates.

[☆] We would like to thank Ana Rute Cardoso, Regina Riphahn, Owen O'Donnell, Marianne Simonsen, participants of the European Association for Labour Economists and the workshop on "health and the labor market", at Molskroen, Denmark, and seminar participants at Antwerp University, the Paris School of Economics for useful comments.

* Corresponding author at: Department of Economics, VU University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands.

E-mail addresses: m.lindeboom@vu.nl (M. Lindeboom), ana.llana-nozal@oecd.org (A. Llana-Nozal), b.vander.klaauw@vu.nl (B. van der Klaauw).

This paper focuses on the causal effect of the onset of a long standing illness or disability on employment. Employment rates are much lower for disabled workers. For instance, according to the UK Department for Work and Pensions in 2012, 46.3% of the working age disabled were employed, while this was 76.4% for the non-disabled. Similar employment gaps exist for other countries (e.g. Von Gaudecker et al., 2011, for the Netherlands). This large employment gap is of direct importance for policies that intend to prevent the onset of disabilities and to increase the labor market prospects of people with disabilities, such as the UK's Disability Discrimination Act (DDA) or the Americans with Disability Act (ADA). Deleire (2000), Acemoglu and Angrist (2001) and Hotchkiss (2003) study the employment effects of the ADA and Bell and Heitmuller (2005) the employment effects of DDA. If the size of the causal effect of disability on employment is large, then the potentials for programs aiming to reduce the impact of disability on employment are large. DDA and ADA are examples of such programs. Besides disability, an individual's employment status strongly depends on demographic and socioeconomic factors. Therefore, for policy purposes it is important to know how much of the difference in

employment rates of disabled and non-disabled workers can be ascribed to background characteristics and how much to the direct causal effect of disability.¹

The issue is also of interest for the large literature on the association between socioeconomic status and health (see e.g. Smith, 1999, for a survey). Even though many economic and epidemiological studies have established a strong positive association ('gradient') between health and socioeconomic status, little is known about causal mechanisms. Assessing causal relations with observational data is non-trivial. Not only may there be direct effects of health (disability) on socioeconomic status and the other way around, but also unobserved individual specific effects can relate to both health and work. Therefore, independent variation in health is required to assess its causal effect on work.

We use accidents that have caused a visit to a hospital, an outpatient facility or a casualty department as a measure for unanticipated health shocks. These accidents include, for example traffic accidents, work place accidents, heart attacks and sport injuries. Such health shocks contain new information to the individual and thereby provide some unanticipated variation. We use this information to identify our model for health shocks, disability and work and subsequently use the model to simulate the causal effect of disability on employment and on the importance of this causal effect in explaining the employment gap that exists between the disabled and the non-disabled model. The unanticipated nature of the shocks is important for the identification of the model, but in our context this only means that the exact timing of the accident is not known in advance. It does not rule out that individuals may be aware that at some moments the risk of experiencing an accident may be higher than in other periods, for instance because this risk depends on current employment status. Also we do not require accidents to be exogenous, conditional on observable characteristics. To be a bit more specific about the model, we construct a discrete-time discrete-choice model for accidents, disability and work. The transition rates between disability and work states can be affected by accidents and accidents can in turn be influenced by the individual's employment and disability status. The three endogenous variables of the model are related via unobserved components that remain fixed across time. Identification of this kind of models has been discussed extensively by Abbring and Van den Berg (2003) and Heckman and Navarro (2007).

Our approach relates to a growing literature that uses (natural) experiments to identify causal relations between socioeconomic status and health. For instance, Lindahl (2005) used lottery prize winners and Snyder and Evans (2006) used changes in the social security law to assess the causal effect of income on health. They find small effects of income on health. This is in agreement with Case and Deaton (2005) and Smith (1998), who conclude that the larger part of the association between health and socioeconomic status at middle and older ages is driven by an effect of health on socioeconomic status, rather than the other way around. Møller-Danø (2005) uses a propensity score and a difference-in-difference matching method to estimate the causal effect of road injuries on income and employment. She finds short and long-run effects of road accidents on employment status for men, but not for women. Lechner and Vasquez-Alvarez (2011), García Gómez and López Nicolás (2006) and Garcia-Gomez et al. (2013) use matching methods to identify the effects of work limitations on employment and income. These papers find significant negative effects of health on employment and income. There are some studies in development economics that use field experiments to assess the effects of interventions affecting health on socioeconomic outcomes. For instance, Miguel and Kremer (2004) evaluate a program of a school-based treatment with a deworming drug in Kenya.

We estimate our model using data from the British National Child Development Study (NCDS). The NCDS is a longitudinal study of around 17,000 individuals born in Great Britain in the week of 3–9 March 1958. These individuals are followed from birth up to the year 2000, when they were 42 years old. At age 40 already about 12% of the respondents face a permanent disability and about 29% of these disabled are out of work. In the full sample the association between disability and employment (the employment gap) is almost 23%-points. Our results show that a health shock causally increases the probability of the onset of a disability with 172%. However, because the health shocks are rare events, the larger part of the onset of disabilities come from a gradual deterioration in health. Furthermore, we find that health shocks affect an individual's labor market status only indirectly through the onset of a disability. Model simulations show that the causal effect of the onset of a disability at age 25 on the employment rate at age 40 is -0.144 . We find large differences between males and females and high and low educated workers. Male employment rates at age 40 are about 23%-points reduced due to a disability, while for females this is 12%-points. Employment rates at age 40 of low educated workers are reduced with 21%-points, while for high educated workers this is only 9%-points. We show that in the complete sample about two-third of the association between disability and employment can be explained by the causal effect of the onset of a disability on employment. The remaining one-third is selection. However, for women selection is more important in explaining the association, while for men and lower educated workers the association is mainly explained by the causal effect from disability to work.

The structure of the paper is as follows. Section 2 discusses the theoretical background and the empirical model. Section 3 introduces the NCDS data and reports on the variables used in the empirical part. Empirical results are discussed in Section 4. Section 5 concludes.

2. Theoretical background and the empirical model

Health production models (Grossman, 1972) or related models (e.g. Cropper, 1977; Ehrlich and Chuma, 1990; Sickles and Yazbeck, 1998; Case and Deaton, 2005) assume that individuals inherit an initial stock of health, which depreciates with age and increases with health investments. Individuals are rational agents who include expectations about their health when making health investments (such as health care consumption and work). If health trajectories are predictable, individuals anticipate that and change their behavior accordingly. So an observed change in labor market status that precedes a health transition can be the result of anticipated behavior rather than labor market status causally affecting health. An unforeseen shock contains new information to the individual and thereby provides some unanticipated variation in health that is unrelated to work status.

In this paper, we consider accidents as unanticipated health shocks providing new information to the individual. We will be more specific about the definition of accidents in the next section when we discuss the data. Such a health shock may cause the onset of a permanent disability or chronic condition. Here our approach differs from, for example, Smith (2003) and Adams et al. (2003) who use the onset of a chronic condition as a measure for health shocks. Health shocks occur at different moments in life and, therefore, our model should be dynamic. A dynamic model also has the advantage that we can substantially relax the requirements for accidents to be valid health shocks. Within our dynamic model we do not restrict health shocks to be exogenous. Instead we explicitly model the occurrence of a health shock and allow unobservables to affect jointly the probability of experiencing a health shock, the onset of disabilities and labor market outcomes. No anticipation of health shocks means that people cannot fully predict the exact timing of the occurrence of such a shock. An intuitive justification of this assumption is that if people would know the exact timing of having an accident in advance, it would be easy to circumvent having

¹ Like most studies in the field we focus on self-reported disability measuring restrictions in daily life (i.e. Berthoult, 2006; Currie and Madrian, 1999). This avoids a mechanical effect which disability insurance benefits receipt has on employment.

the accident.² This allows identifying the causal effects of the health shock on the disability and employment status without exclusion restrictions or strong functional form restrictions. Abbring and Van den Berg (2003) and Heckman and Navarro (2007) provide extensive discussions on the identification of causal effects in event-history models and discrete-time discrete-choice models, respectively. We return to identification issues later.

2.1. Empirical specification

The data contain individuals who were all born in the week of 3–9 March 1958. These individuals are followed from birth up to age 42. We have constructed individual histories from the moment the individual leaves full-time education up to the end of the observation period. For about 60% of the individuals we observe the complete history. We assume that the attrition of the other individuals is exogenous (see Section 3 for a justification of this assumption). The individual histories contain yearly information on employment status and health status. Let $S_l(t)$ denote the individual's labor market status on his/her t th birthday, which can either be working (1) or non-working (0). Since we only start following individuals after leaving full-time education, non-working does not include full-time education. The variable $S_h(t)$ denotes the health status on the individual's t th birthday, which can either be disabled (1) or non-disabled (0). The focus of the paper is on permanent disabilities (and thus we ignore short-term limitations), which makes disability an absorbing state. The health shock is denoted by the variable $A(t)$ which equals 1 if an individual experienced a shock at age t and 0 otherwise.

Our key variables $S_l(t)$, $S_h(t)$ and $A(t)$ are observed yearly from the moment that an individual leaves full-time education at age T_0 and remains participating in the surveys until age T_1 . So, for each individual we observe the occurrence of shocks and the transitions between four different states: (1) employed/non-disabled, (2) non-employed/non-disabled, (3) employed/disabled and (4) non-employed/disabled. Because disability is an absorbing state, transitions from (3) and (4) to (1) and (2) are ruled out. There are thus eight transitions possible. Since the data are observed at specific moments in time, we use a discrete-time discrete-choice model rather than, for example, a continuous-time duration model. The main advantage is that the model has the same time unit as the data, which avoids computational difficulties. The disadvantage of a discrete-time model is that events occur in real time and that we ignore possible dynamics within the time period. In our model we allow annual transition probabilities between states to depend on current and past health shocks. Below we sketch the setup of our model, while details are provided in Appendix A.

Health shocks do not occur exogenously. For example, those at work may be exposed to adverse working conditions, which may increase the risk of experiencing a health shock. Therefore, we allow the probability of the occurrence of a health shock in time interval $<t, t+1]$ to depend on the individual's labor market and disability status at time t and whether or not the individual experienced a health shock in the previous period. In addition we allow the health shock probability to depend on observed characteristics at time t ($x(t)$) and a time-invariant unobserved component v_a . The vector of observed individual characteristics $x(t)$ includes an intercept, a set of socioeconomics characteristics and a fourth order polynomial in age. Recall that because all individuals in the data are born in the same week, the polynomial in age also includes calendar time effects such as business cycle fluctuations. We use a logit specification to model the probability of an accident at age t .

Health shocks can affect the probability that individuals move between different disability and employment states in two ways. First, there is an instantaneous effect, and second, we allow health shocks to

have a permanent long-run effect.³ Once an individual has experienced a health shock, this might affect his/her transition probabilities in all following years. Again we use logit specifications for the transition probabilities between the different disability and work states and these logit models depend again on observed characteristics at time t ($x(t)$) and time-invariant unobserved components (v).

Because we use a separate logit specification for each possible transition, the impact of a health shock can be different for individuals in different employment and disability states. Health shocks are not only related to the transition probabilities through these effects, but we also allow for interdependence through the time-invariant unobserved heterogeneity components. The unobserved component in the health-shock probability v_a may be related to the unobserved components v in the transitions probabilities. The latter accounts for endogeneity of the health shocks. We use a random effects specification for the unobserved heterogeneity, and in particular a so-called factor-loading specification to allow for correlations between the different unobserved heterogeneity terms (see Appendix A for details). Since the model is fully parameterized, we use maximum likelihood to estimate the parameters. Appendix A shows the likelihood function.

The main parameters of interest describe the effects of health shocks on disability and work outcomes. A health shock is defined in our data by a question whether an individual had "been admitted to a hospital or attended a hospital outpatient or casualty department as a result of any kind of accident". To the individual, a health shock is an event which cannot be anticipated. No anticipation in our context means that transition probabilities prior to the occurrence of a health shock are not affected by a future health shock and also not by the moment at which this health shock will occur. For example, conditional on observables and unobservables the transition probabilities at age 25 are the same if this individual experiences a health shock at age 26 or at age 40. The health shock thus does not have any impact on transition probabilities prior to the moment at which the individual has a health shock. This does not imply that health shocks are exogenous or that each individual has in each time period the same probability of a health shock. It rather implies that individuals cannot foresee the exact moment at which a health shock occurs to them. So individuals might know that in some periods the probability of a health shock is high, for example when they are employed or as they get older.

The assumption that individuals do not anticipate the exact moment of arrival of the health shock plays a crucial role in the identification of event history models. From the definition of health shocks (see above), it is clear that this measure satisfies the condition that the event cannot be anticipated. We refer to Appendix A for a more elaborate, but intuitive explanation of the identification of our model. Abbring and Van den Berg (2003) provide a formal discussion of identification in event history models.

However, a few issues are relevant. First, there might be selectivity in the decision to go to hospital after the occurrence of an accident. It is important to note that the UK has a National Health Care system and that in principle health care is accessible for all.⁴ Of course, it remains an individual decision whether or not to go to hospital after an accident. For example, an individual's socioeconomic status may affect the probability that the individual visits the hospital after having experienced an accident. Such differences are controlled for by the observed characteristics and unobserved characteristics in the model for accidents. This might also imply that the effect of a health shock is not the same for all individuals. Therefore, we allow the health shock effect to depend on both observables and unobservables. We return to this issue below.

³ We also tried including medium-run effects of health shocks and more flexible specification, where the effect of a health shock depends on the elapsed duration since the occurrence of the shock. But this did not substantially improve the model.

⁴ It is possible to buy supplemental private health insurance. This private insurance allows individuals to have more freedom in the choice of provider, to have a private room, when hospitalized and to have shorter waiting times for specialized care. Only about 15% of the individuals have private insurance.

² We used a fixed effects panel data model to investigate if accidents are related to other events such as graduation, but we could not find any strong evidence for this.

Second, our definition of a health shock (see the next section) is that it is an accident that leads to a hospital, outpatient or casualty department visit. Accidents that did not directly lead to hospital visits are thus included in the gradual deterioration of health. This also holds for health events that may have caused hospital care in the longer run. The distinction between gradual and sudden changes of health, therefore, strongly depends on the definition of the health shock variable. Of prime importance for our empirical analysis is the causal effect of the onset of a disability on employment and for this we require random variation in the timing of health shocks. With this strict definition of a health shock we ensure that this condition is satisfied.

Third, the size of the impact of a disability on employment is for an important part determined by social insurances, and in particular disability insurance programs. Disability insurance is, like any other insurance program, subject to moral hazard and, therefore, the eligibility conditions and benefit level is important for exit rates out of employment into disability insurance. For an extensive overview of the UK disability insurance program over the years we refer to Banks et al. (2012). Of relevance here is that UK disability insurance program is characterized by a flat rate (i.e. it is independent of previous earnings) that is only about 15% of average income, making it one of the least generous programs in developed countries. Also, over the period considered, some changes in the program have been implemented. We are following a single cohort over time and, therefore, in our model changes over time will be picked up by the age variables.

Fourth, the identification results apply in our model to the effects of health shocks on the disability and employment status. We are ultimately interested in the causal effect of the onset of a disability on employment. For this we need that health shocks do not have direct effects on employment if the individual's health status remains unchanged. We do not impose this assumption on our model, but our empirical results (shown in Section 4), do indeed show that this is the case. The health shock causes some variation in the disability status, which in turn is used to identify the causal effect of the onset of a disability on employment status. We exploit this: after estimation of the model we simulate the model and derive estimates of the causal effect of the onset of a disability on later employment outcomes (details of the simulations are provided in Appendix C).

Besides this baseline specification of our model we also have estimated more elaborate models with heterogeneous health shock effects, i.e. models where health shocks have different effects on males, or where the impact may also vary with age and educational attainment and even where the effects depend on the unobservables. None of the parameters describing the differential effects of health shocks by gender, age and education were individually significant, making it difficult to draw strong conclusions (results available upon request). We, therefore, did not pursue this issue further.

3. Data

3.1. Sample

We use the National Child Development Study (NCDS), a longitudinal study that follows about 17,000 individuals born in Great Britain in the week of 3–9 March 1958. The study started as the “Perinatal Mortality Survey” and surveyed the economic and obstetric factors associated with stillbirth and infant mortality. Since the first survey in 1958, cohort members have been traced on six other occasions to monitor their physical, educational and social circumstances. The waves were carried out in 1965 (age 7), 1969 (age 11), 1974 (age 16), 1981 (age 23), 1991 (age 33) and 1999/2000 (age 42). In addition to the main surveys, information about the public examinations was obtained from the schools in 1978. For the birth survey, information was gathered from the mother and the medical records. The interviews for wave 1 to 3 (covering age 7, 11 and 16) were carried out with parents, teachers, and the school health service; while ability tests were administered to the

cohort members. The subsequent surveys included information on employment and income, health and health behavior, citizenship and values, relationships, parenting and housing, education and training of the respondents. In waves 4, 5 and 6, individuals were asked to retrospectively give information on their employment, unemployment, out-of-the-labor-force and education/training periods, recording their starting and ending dates. The NCDS is, therefore, highly appropriate to look at life histories.

Our model describes transitions between disability and labor market states from the moment that individuals leave full-time education. We use a sample of 12,375 individuals who participated in the 1981-survey at age 23 and for who we can construct labor market, health and health shock histories.⁵ Case et al. (2005) investigate attrition from the survey by comparing low birth weight and father's occupation across different NCDS waves. They do not find any evidence for non-random attrition with respect to these variables. Furthermore, advisory and user support groups of the NCDS compared respondents and non-respondents in the later surveys in terms of social and economic status, education, health, housing and demographics. It is found that the distribution of these variables among the sample survivors do not differ from the original sample to any great extent (NCDS User Support, 1991). In addition, the 1981 sample is compared to the UK 1981 Population Censuses in terms of the distributions of key variables such as marital status, gender, economic activity, gross weekly pay, tenure and ethnicity (Ades, 1983). The overall conclusion is that the sample appears to be representative with respect to these variables.

3.2. Dependent variables: labor market status, disability status and health shock

The labor market status is measured each year in March, around the time of the birthday of the individuals. We distinguish two labor market outcomes, employed and non-employed. An individual is considered to be employed if either the individual has a full-time or part-time job, is self-employed or on maternity leave. Also an apprenticeship scheme which is part of a job is considered as employment. Currie and Hyson (1999), who use the same data to examine the long-run effect of low birth weight on educational attainment and labor market outcomes, find that their empirical results are not sensitive to the exact definition of employment. In Figs. 1 and 2, we show the employment rate, the unemployment rate and the fraction of individuals out of the labor force and in full-time education at different ages for both men and women. The figures show that employment rates are higher for males, that a substantial share of the females is out of the labor force and that this fraction declines with age. For males the fraction of workers out of the labor force increases gradually with age. The unemployment rates are higher between the ages of 22–25, these ages coincide with a period when the UK experienced a severe recession.

We define a disability/longstanding illness from the response to the following question:

Do you have any longstanding illness, disability or infirmity of any kind?

The respondents are asked to report whether they had longstanding illnesses, disabilities and infirmities for every year at around the time of their birthday. This implies that our data are not fully informative on short-term limitations, which might have lasted a few months between two birthdays. If a disability is reported, it is asked from which conditions the respondent suffered. We only consider conditions that are included in the International Classification of Diseases (ICD-9) produced by the World Health Organization (1977). The ICD-9 is extensively used in epidemiological and health management studies to classify diseases and health problems (World Health Organization, 2004). These

⁵ 60% of the individuals in our sample are present in wave 4 (age 23), 5 (age 33) and 6 (age 42), 28% only in wave 4 and 12% in waves 4 and 5.



Fig. 1. Labor market states of males.

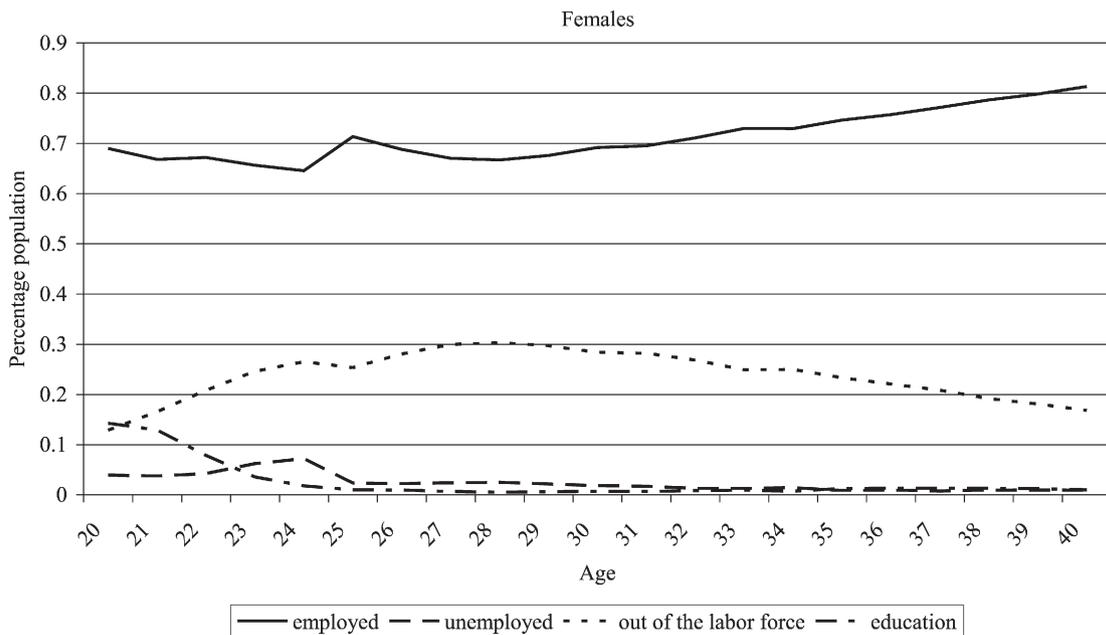


Fig. 2. Labor market states of females.

include, for instance, serious disabilities that limit work/daily activities such as epilepsy, blindness, deafness, multiple sclerosis, mental retardation, a congenital condition, or a traumatic amputation or internal injury (see Appendix B for the complete list). We only consider an individual as disabled if the individual reports having the same longstanding illness, disability or infirmity for at least three consecutive years.^{6,7}

⁶ We add this condition to be sure that we have long lasting impairments. For instance, if an individual reports to have complications of pregnancy (code 11 in the ICD-9 classification) during one year, then the individual is not counted as permanently disabled. Disabilities listed in the ICD-9 code lasting more than three years are considered to be permanent. The data confirm that recovery rates after three years are very low.

⁷ We follow our respondents up to age 42. The extra condition that the impairment should last at least three years implies, therefore, that in the estimation of the model we only use the disability and work history information up to the age of 40.

Indeed, the data confirm that recovery rates after three years are negligible. Case et al. (2005), use responses to the question "How good is your health" and find that these self reports are strongly correlated to the question on the prevalence of long standing illnesses, disabilities and infirmities. Bajekal et al. (2004) show in a report commissioned by the UK Department for Work and Pensions that age-specific disability rates for employed workers do not vary much across surveys using different definitions for disabilities.

Fig. 3 shows the fraction of individuals with a disability after age 16. Disability rates are very similar for men and women. At age 20 over 4% of the individuals in the sample has some disability. This increases up to about 12% at age 40. Some people already have long standing disabilities that started during childhood, but the majority of the disabilities started during working ages. In fact, the slope becomes somewhat steeper at

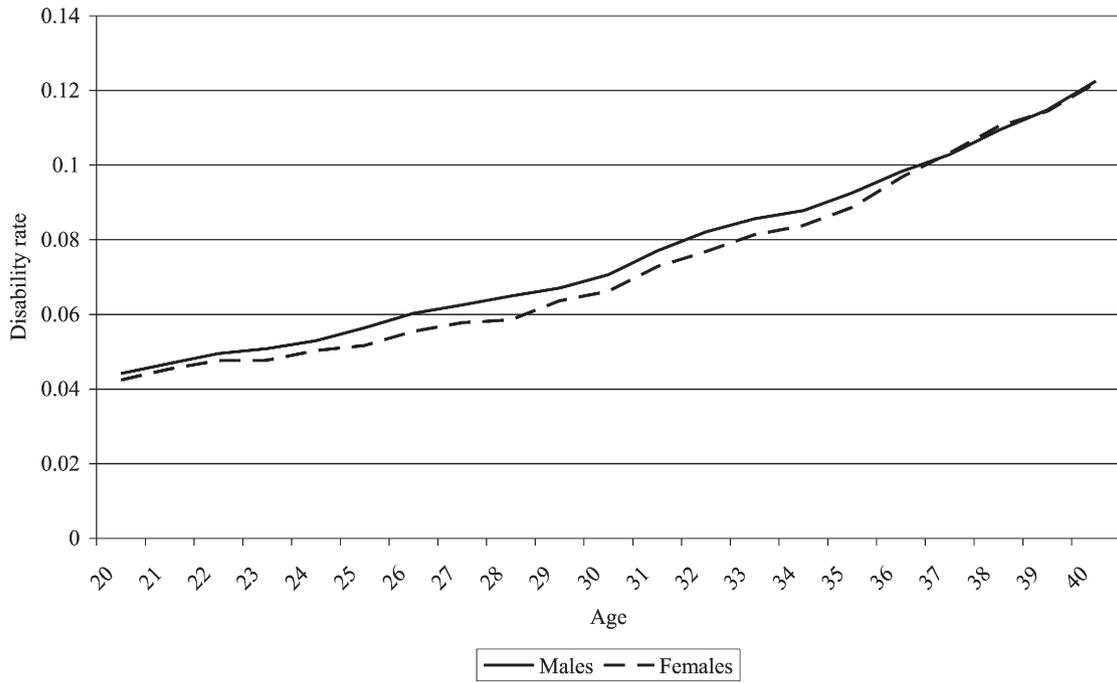


Fig. 3. Disability rates of males and females.

older ages, which means that the hazard of the onset of a disability becomes larger as people get older.

We derive our measure for a health shock from the following question:

Have you been admitted to hospital or attended a hospital outpatient or casualty department as a result of any kind of accident?

We observe both the date of the health shock and the cause.⁸ Men are much more likely to experience a health shock than women. In our sample, around 77% of the men had at least one health shock during the observation period, while this was only about 42% for women. We follow individuals for more than two decades and, therefore, quite a few respondents have experienced more than one health shock during this period. Not only the incidence of health shock differs between men and women, but also the cause differs. Table 1 lists health shocks by cause and gender. For each cause men are much more likely to experience a health shock than women. The most substantial difference in incidence rates occurs for work and sports-related accidents. Because of this large share of work-related accidents, it is particularly important in our empirical analyses to allow the health shock probability to depend on labor market status. Fig. 4 plots annual incidence rates by age and shows that for both men and women the probability of a health shock is relatively high until the mid-twenties and drops substantially afterwards.

We use the annual labor market status and disability status to classify each individual in each year in one of four states: (1) work and disabled, (2) non-work and disabled, (3) work and non-disabled and (4) non-work and non-disabled. In Fig. 5 we show for different ages the fraction of individuals in each state. At all ages most individuals are employed and non-disabled. At later ages the fraction of individuals in the non-work non-disabled state decreases while the fractions of individuals increase in both disabled states (either with or without work). The figure shows a fall in employment rate at ages 22–25, this is in line with the observed patterns in Figs. 1 and 2. Our empirical model is

specified in terms of yearly transition probabilities between the different states. Table 2 provides a summary of the yearly transitions for both men and women. The table shows that there is a high degree of serial correlation and that individuals are much more likely to change labor market status than disability status. The table also shows that 16.8% of the men with a disability find a job in the subsequent period. This means that there remain employment opportunities for people with a disability. However, note that this number should be contrasted with the 41.9% of non-disabled workers who find a job in the next period. For females the differences between the job-finding rate of disabled and non-disabled workers is much smaller (12.8% versus 19.3%). Of course, these differences reflect associations and cannot be given any causal interpretation. Furthermore, we cannot distinguish between differences in which disabled and non-disabled workers search for work and the probability of finding work given a particular search level.

3.3. Background variables

The role of (early) childhood conditions in explaining later life health and socio-economic status is well documented (see, for instance, the literature reviewed in Case et al., 2005; Almond and Currie, 2011). The NCDS is very rich on individual characteristics and particularly on childhood characteristics. For each individual we observe a range of variables that give information on an individual's birth weight and health, cognitive ability and socioeconomic background in childhood. In constructing the relevant background variables we follow the definitions used by Case et al. (2005) and Currie and Hyson (1999). Table 3 provides sample means on these variables. For many variables there is

Table 1
Yearly incidences of different types of accidents.

	Male	Female
Overall	0.1199	0.0391
Road (pedestrian)	0.0018	0.0013
Road (driver)	0.0179	0.0080
Workplace	0.0398	0.0072
Home	0.0127	0.0107
Sports	0.0338	0.0047
Other	0.0139	0.0072

⁸ The questionnaire restricts the number of accidents that can be reported to 8 in the 1981-wave and 6 in the 1991 and 1999/2000-wave. In each wave only between 1 and 2% of the individuals actually reports this maximum.

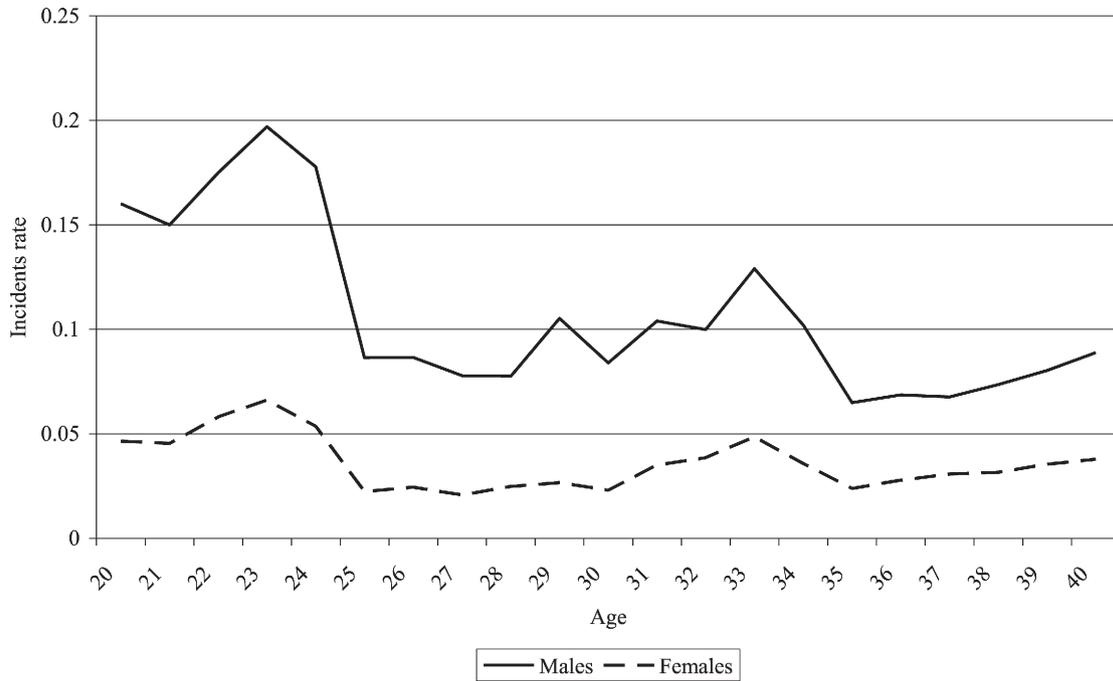


Fig. 4. The annual incidence rates of health shocks for males and females.

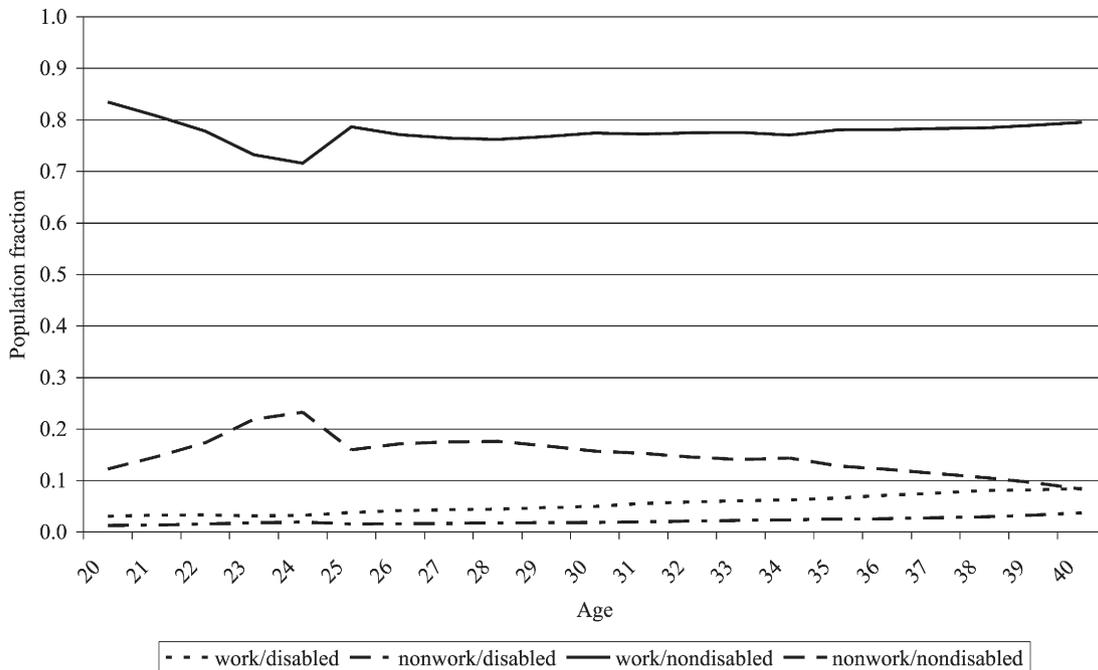


Fig. 5. Disability and employment states.

some item non-response. To avoid losing many observations we follow Case et al. (2005) by constructing dummy variables that indicate if the information on a variable is missing. So if an individual's characteristics are missing, then the variables describing these characteristics take value 0 while the dummy variable indicating that these variables are missing takes value 1.⁹

⁹ The alternative is either to leave covariates which are missing for some individuals out of the model or to drop individuals with missing observations from the data. The role of covariates in our model is to account for as much individual variation as possible. Covariate effects should therefore also be interpreted as associations rather than as causal effects.

For each individual we observe the educational attainment. The level of education is derived by compiling an education variable with categories aggregated to national vocational qualification levels. We include the following categories: less than O-level, O-level equivalent, A-level equivalent and degree equivalent. In general, O-level exams are taken at age 15/16 and A-level exams two years later. Degree equivalent implies that the individual has additional education on top of secondary education.

The family's socioeconomic status is derived from the father's social class at birth. The social class corresponds to a system used by the British Registrar General and consists of: professional, supervisory, skilled non-

Table 2
Transition matrices between work and disability states by gender.

State in year t	State in year t + 1			
	Work/ disabled	Nonwork/ disabled	Work/ nondisabled	Nonwork/ nondisabled
<i>Male</i>				
Work/disabled	95.3%	4.7%		
Nonwork/disabled	16.8%	83.2%		
Work/nondisabled	0.3%	0.1%	96.8%	2.8%
Nonwork/nondisabled	0.3%	0.7%	41.9%	57.2%
<i>Female</i>				
Work/disabled	90.3%	9.7%		
Nonwork/disabled	12.8%	87.2%		
Work/nondisabled	0.3%	0.0%	91.7%	7.9%
Nonwork/nondisabled	0.1%	0.4%	19.3%	80.2%

manual, skilled manual, semi-skilled non-manual, semi-skilled manual, and unskilled. We classify socioeconomic status as high if the father is in a professional, supervisory, skilled non-manual job; medium if the father is in skilled manual, semi-skilled non-manual; and low if the father is in a semi-skilled manual and unskilled job. Following Currie and Thomas (1999), we classify individuals whose father's information is missing by the mother's social class. In case the social classes of both parents are missing, we assign the individual to low socioeconomic status if the mother was single and to missing if both parents were present.

Low birth weight is a dummy variable for infants with a birth weight below 2500 g. There is evidence from the epidemiological literature that low birth weight is strongly associated with infant and later life mortality (World Health Organization, 2004). Low weight at birth can be the result of either preterm birth (before 37 weeks of gestation) or restricted fetal growth. In the empirical analyses we do not make a distinction between these two categories. We create a dummy variable that indicates if the mother smoked after the fourth month of pregnancy. Smoking during pregnancy has been found to be related with cognitive deficiencies and other health problems in the medical and epidemiological literature (see for instance Blair et al., 1996; Conter et al., 1995; Naeye and Peters, 1984; Williams et al., 1998). Furthermore, we observe the mother's age at birth. Mother's age at the child's birth can influence the child's health through, for instance nutritional

Table 3
Sample mean of the individual characteristics.

	Total	Male	Female
Female	50.2%		
Education (national vocational qualification level)			
Below O-level equivalent	26.6%	24.6%	27.7%
O-level equivalent	31.5%	27.9%	35.1%
A-level equivalent	17.1%	20.9%	13.3%
Degree equivalent	25.2%	26.6%	23.8%
Parental information at birth missing	7.3%	7.6%	7.1%
High parental socioeconomic status at birth	25.2%	25.5%	24.9%
Medium parental socioeconomic status at birth	46.7%	46.0%	47.4%
Low parental socioeconomic status at birth	20.8%	20.9%	20.6%
Mother smoked after the fourth month of pregnancy	30.5%	30.0%	31.0%
Mother did not smoke after the fourth month of pregnancy	62.2%	62.4%	61.9%
Birth information missing	5.5%	5.8%	5.3%
Mother's age at birth (in years)	27.6	27.5	27.6
Low birth weight (less than 2500 g)	4.8%	4.1%	5.4%
Normal birth weight (more than 2500 g)	89.7%	90.1%	89.3%
Born in North	27.1%	26.5%	27.7%
Born in Midlands	23.5%	24.3%	22.6%
Born in South & Wales	16.3%	16.1%	16.5%
Born in Scotland	10.4%	10.0%	10.7%
Born in London & South-East	17.3%	17.3%	17.2%
Test scores at age 7 missing	12.0%	12.6%	11.5%
Math	5.2	5.3	5.1
Bristol Social Adjustment Guide	8.4	9.8	7.0
Number of observations	12,375		

Table 4a
Parameter estimates of the logit model for health shock probabilities.

	Parameter estimates
Intercept	0.987 (0.013)
Being employed	0.342 (0.014)
Being disabled	0.048 (0.004)
Female	−1.203 (0.010)
Age (divided by 40)	−2.547 (0.006)
(Age/40) ²	−1.190 (0.007)
(Age/40) ³	0.094 (0.007)
(Age/40) ⁴	1.493 (0.008)
O-level equivalent education	0.041 (0.009)
A-level equivalent education	0.171 (0.011)
Degree equivalent education	0.041 (0.011)
Childhood background variables	
Low socioeconomic status at birth	−0.035 (0.007)
Mother smoked at pregnancy	0.076 (0.005)
Birth information missing	−0.730 (0.009)
Mother's age at birth (divided by 10)	−0.290 (0.008)
Mother's age at birth squared (divided by 100)	0.423 (0.014)
Low birth weight	−0.016 (0.004)
Born in London or South-East	0.055 (0.005)
Born in Scotland	−0.103 (0.004)
Born in South or Wales	0.007 (0.005)
Born in North	0.043 (0.005)
Born in Midlands	0
Scores at age 7 missing	0.066 (0.008)
Math score (divided by 10)	0.156 (0.008)
Bristol Social Adjustment Guide (divided by 10)	0.051 (0.005)
Parameters of the mixing distribution	
Probability 1: $\theta_1\theta_2$	0.129 (0.002)
Probability 2: $(1 - \theta_1)\theta_2$	0.423 (0.005)
Probability 3: $\theta_1(1 - \theta_2)$	0.105 (0.001)
Probability 4: $(1 - \theta_1)(1 - \theta_2)$	0.344 (0.004)
Location mass point 1	0
Location mass point 2	−0.998 (0.010)
Location mass point 3	1.204 (0.013)
Location mass point 4	0.206 (0.019)

Standard errors in parentheses.

See Table 4b for the number of observations and the value of the log likelihood function.

deficiencies if the mother is very young, or delivery complications if the mother is older. Therefore, in the empirical analyses we include also the mother's age at birth squared. We use the region at birth to control for geographical differences and/or differences in labor market conditions.

For each individual we observe test scores on math and social adjustment at age 7. The math test is designed for the NCDS and assesses arithmetic ability. The score ranges from 0 to 10. Currie and Thomas (1999) show that test scores at the age of 7 have significant impacts on later education attainments and labor market outcomes. The Bristol Social Adjustment Guide is designed to assess the child's social behavior in school and at home. The test is completed by the teacher who knows the child best. Higher scores indicate higher maladjustment.

4. Results

4.1. Parameter estimates

Tables 4a and 4b show the parameter estimates of our model. In Fig. 6 we show some indication for the fit of our model. In this figure we compare the observed data to model predictions.¹⁰ The model predictions manage to pick up the general life cycle trends in employment and disability. Only around age 23/24 the model has difficulties with matching the fluctuations in observed employment rates.

We start with a brief discussion of the results for the mixing distribution (unobserved heterogeneity). The parameters of this distribution

¹⁰ Our model describes transitions between health and labor market states after having left full-time education. We have estimated a multinomial logit model for the first state after leaving full-time education.

Table 4b
Multinomial logit with unobserved heterogeneity on transitions between work and disability states.

From	Disabled		Nondisabled					
	Work	Nonwork	Work		Nondisabled		Nonwork	
To	Disabled		Disabled		Nondisabled		Nondisabled	
	Nonwork	Work	Work	Nonwork	Nonwork	Work	Nonwork	Work
Intercept	-2.491 (0.010)	-1.213 (0.004)	-6.185 (0.005)	-9.216 (0.021)	-4.166 (0.014)	-3.411 (0.004)	-4.023 (0.006)	2.241 (0.025)
Instantaneous effect health shock	-0.074 (0.004)	0.462 (0.003)	0.862 (0.004)	1.658 (0.006)	-0.019 (0.010)	1.021 (0.005)	0.863 (0.004)	0.045 (0.033)
Permanent effect of health shock	-0.003 (0.010)	0.474 (0.006)	0.159 (0.004)	0.409 (0.005)	-0.059 (0.007)	-0.322 (0.003)	0.116 (0.005)	-0.172 (0.027)
Gender	0.922 (0.010)	-0.359 (0.010)	0.315 (0.006)	1.164 (0.005)	1.241 (0.008)	-1.450 (0.006)	-0.635 (0.005)	-1.290 (0.008)
Age (divided by 40)	-0.116 (0.004)	0.358 (0.004)	0.245 (0.006)	0.062 (0.003)	3.668 (0.035)	-0.128 (0.005)	-0.212 (0.004)	-2.595 (0.021)
(Age/40) ²	-0.533 (0.005)	-0.408 (0.004)	0.588 (0.005)	0.101 (0.004)	0.039 (0.009)	-0.113 (0.004)	0.302 (0.004)	-1.295 (0.006)
(Age/40) ³	-0.262 (0.004)	-0.450 (0.004)	0.342 (0.003)	0.381 (0.003)	-1.630 (0.011)	-0.055 (0.004)	0.328 (0.003)	0.507 (0.012)
(Age/40) ⁴	0.017 (0.004)	-0.623 (0.004)	0.219 (0.004)	0.676 (0.006)	-2.429 (0.010)	-0.035 (0.003)	0.543 (0.005)	1.513 (0.009)
O-level equivalent education	-0.265 (0.007)	-0.659 (0.007)	-0.691 (0.004)	0.666 (0.004)	0.683 (0.010)	0.869 (0.007)	-0.216 (0.004)	-0.272 (0.004)
A-level equivalent education	-0.523 (0.005)	-0.468 (0.009)	-1.121 (0.018)	-0.857 (0.016)	-0.521 (0.013)	-0.789 (0.011)	-0.913 (0.014)	0.193 (0.003)
Degree equivalent education	-0.460 (0.010)	-0.009 (0.004)	-0.171 (0.003)	0.090 (0.004)	0.291 (0.004)	0.430 (0.013)	0.524 (0.011)	0.774 (0.012)
Parental socioeconomic status at birth								
Missing	0.450 (0.003)	-0.080 (0.005)	0.448 (0.005)	0.719 (0.011)	0.347 (0.013)	-0.041 (0.003)	-0.850 (0.007)	-0.115 (0.010)
High	-0.099 (0.003)	-0.316 (0.004)	-0.085 (0.004)	-0.299 (0.004)	-0.056 (0.005)	0.314 (0.004)	-0.152 (0.003)	0.037 (0.007)
Low	0.111 (0.005)	-0.028 (0.005)	0.147 (0.003)	0.132 (0.003)	0.183 (0.005)	0.576 (0.003)	-0.235 (0.003)	-0.125 (0.010)
Mother smoking at pregnancy	0.123 (0.003)	0.007 (0.004)	0.159 (0.004)	0.358 (0.003)	0.122 (0.010)	-0.315 (0.003)	0.184 (0.003)	0.006 (0.009)
Birth info missing	0.551 (0.003)	-0.069 (0.003)	-0.514 (0.004)	-0.559 (0.006)	-0.571 (0.016)	-0.081 (0.003)	0.090 (0.007)	0.213 (0.009)
Mother's age at birth (/10)	0.193 (0.009)	0.003 (0.009)	-0.098 (0.008)	0.472 (0.007)	-0.252 (0.004)	-0.004 (0.005)	-0.062 (0.003)	0.007 (0.016)
Mother's age at Age (/10) squared	-0.029 (0.007)	-0.129 (0.009)	0.127 (0.007)	-0.691 (0.015)	0.414 (0.006)	0.048 (0.003)	0.111 (0.004)	0.033 (0.022)
Low birth weight	-0.072 (0.005)	-0.323 (0.004)	0.157 (0.003)	-0.042 (0.004)	-0.081 (0.007)	0.044 (0.003)	-0.522 (0.004)	-0.084 (0.007)
Region of residence at birth								
London	0.054 (0.003)	-0.012 (0.003)	-0.121 (0.003)	-0.336 (0.004)	-0.006 (0.004)	-0.471 (0.004)	0.138 (0.003)	-0.024 (0.004)
Scotland	0.217 (0.003)	-0.005 (0.004)	0.044 (0.003)	-0.275 (0.003)	0.236 (0.007)	0.603 (0.005)	0.167 (0.003)	-0.076 (0.004)
South & Wales	0.165 (0.004)	-0.094 (0.003)	0.179 (0.003)	0.117 (0.003)	0.030 (0.003)	0.364 (0.004)	0.070 (0.003)	-0.030 (0.005)
North	0.374 (0.004)	-0.101 (0.003)	-0.056 (0.003)	0.215 (0.003)	0.211 (0.004)	0.217 (0.004)	0.257 (0.003)	-0.055 (0.004)
Score at age 7 missing	-0.145 (0.003)	-0.441 (0.003)	0.161 (0.003)	-0.516 (0.007)	0.156 (0.004)	-0.351 (0.006)	-0.072 (0.007)	0.009 (0.013)
Math score (/10)	-0.387 (0.007)	0.118 (0.006)	-0.043 (0.004)	-0.709 (0.006)	-0.040 (0.009)	0.209 (0.004)	-0.653 (0.008)	0.088 (0.012)
Bristol social adjustment guide (/10)	0.209 (0.004)	-0.185 (0.004)	0.055 (0.004)	0.180 (0.003)	0.140 (0.013)	-0.329 (0.004)	0.101 (0.003)	-0.088 (0.009)
Parameters of the mixing distribution								
Location mass point 1	0	0	0	0	0	0	0	0
Location mass point 2	-1.518 (0.009)	0.443 (0.020)	-0.691 (0.011)	-1.104 (0.014)	-1.365 (0.014)	-0.533 (0.008)	-0.440 (0.010)	0.209 (0.096)
Location mass point 3	-1.003 (0.006)	-0.073 (0.034)	-0.065 (0.005)	-0.962 (0.009)	-0.785 (0.016)	-1.021 (0.008)	0.007 (0.004)	0.983 (0.033)
Location mass point 4	-2.521 (0.007)	0.370 (0.053)	-0.755 (0.012)	-2.066 (0.022)	-2.150 (0.013)	-1.554 (0.004)	-0.433 (0.013)	1.192 (0.088)
Number of observations	12,375							
Value of the -log-likelihood	104,848.05							

are reported in the lower panels of Tables 4a and 4b. Almost all parameters of the mixing distribution are highly significant, indicating the importance to allow for a stochastic relationship between disability, work and health shocks. Recall from Eqs. (3a) and (3b) that we impose a factor-loading specification with two elements that each can take on two values. We have tried to add additional elements to the model, but this increased the computer time for estimating the model substantially while only marginally improving the fit of the model.¹¹ The additional parameters are jointly significant at the 10% level. For each transition we thus have four mass points. The first mass point is normalized to 0 because each model equation includes an intercept. Most probability mass (about 42%) is located at mass point 2, describing individuals with a low probability of experiencing health shocks. Mass point 3 has less probability mass (about 10%) and describes individuals who are most likely to get a health shock. It is not straightforward to detect a clear pattern in the effects of the unobservables in the transition probabilities. But compared to mass point 1, all other mass points describe individuals who are less likely to get disabled and who are more likely to stay in employment or move to employment if the health status remains unchanged.

¹¹ The Akaike information criterion rejects including additional unobserved heterogeneity terms.

Table 4a shows the parameter estimates from the logit specification for the probability of getting a health shock. Employed individuals have about a 41% ($= \exp(0.342) - 1$) higher probability of getting a health shock. Recall from Table 1 that indeed a substantial share of the health shocks is work-related. The health shock rate of an individual with a disability is only about 5% higher than the health shock rate for a non-disabled person. Males are more than three times more likely to get a health shock than females. This is also what we directly observe in the data when comparing males and females (see Table 1). Apparently, differences in background characteristics between males and females cannot explain the differences in health shocks rates. The age coefficients indicate that the probability of getting a health shock decreases up to age 37 and increases afterwards. The socioeconomic characteristics of the parents of the cohort member all have significant impacts on the health shock rate. Individuals with medium socioeconomic status and whose mother smoked during pregnancy have higher health shock rates. The health shock rate also increases with the age of the respondent's mother at birth. Individual characteristics are also important. We find significant effects for birth weight, region and the test scores and education. Most of these effects are in line with a priori expectations. We find a somewhat puzzling effect of the math test score at age 7. Individuals with a higher test score have higher health shock rates. It should be noted that the coefficients purely reflect associations and that one should not connect a strong causal interpretation to these findings.

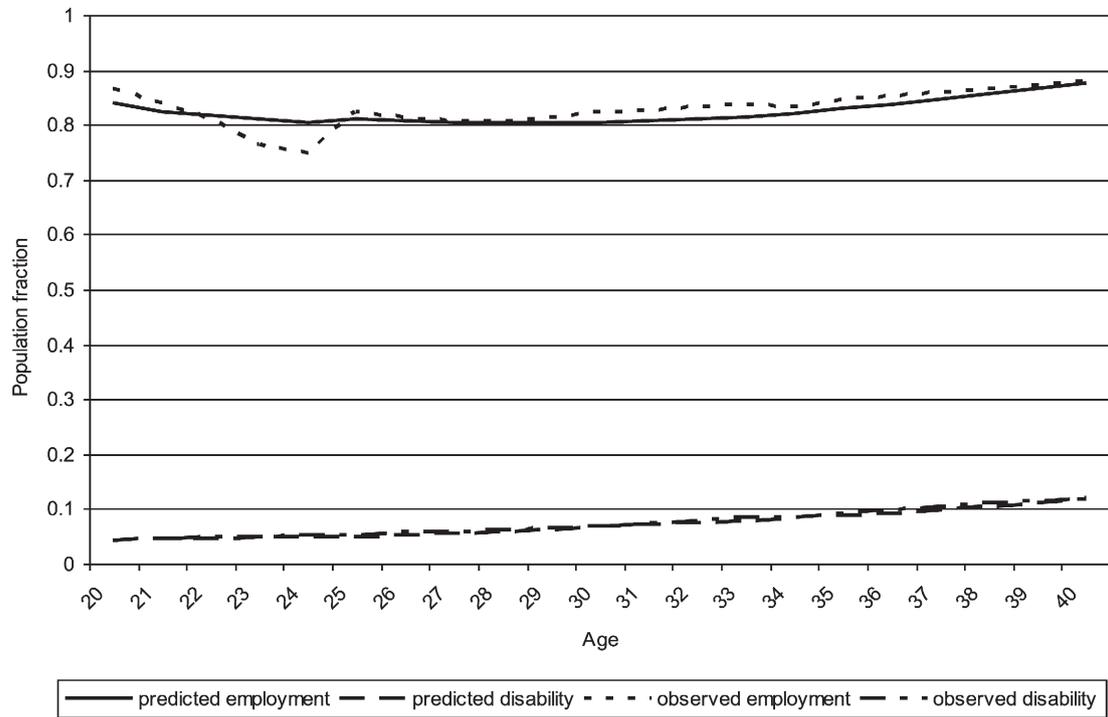


Fig. 6. Fit of the model.

Table 4b shows the parameter estimates of a multinomial logit model for the transitions between different labor market and disability states. The main parameters of interest are the parameters describing the effects of health shocks on transition probabilities. We allow health shocks to have a direct instantaneous effect and a long-run effect. The estimated coefficients for the instantaneous effects are almost always larger than the estimated coefficients for the long-run effects.¹² Furthermore, the instantaneous effects of the health shocks are largest on the transition rates from the non-disabled states to disabled states. The estimated coefficients for transitions between employment states where the disability status remains the same are much smaller. The long-run effects mainly show that in the long-run health shocks not only have an effect on the disability status, but also on the employment status.

4.2. Simulations

Interpreting the coefficients separately in non-linear models with multiple states is difficult. Therefore, to illustrate the impact of a health shock we consider a representative individual and perform calculations with the estimated model.¹³ The details of the simulations are described in Appendix C. Suppose this individual did not experience a health shock until age 23. The probability that this individual is non-disabled at his/her 24th birthday is 0.933 (with standard error 0.00034 computed using the Delta method). Without experiencing a health shock at age 24 the probability of becoming disabled before his/her 25th birthday is 0.0025 (s.e. 0.000062). However, if the individual did experience a health shock at age 24, this probability becomes 0.0069 (s.e. 0.00015). So the health shock increases the instantaneous onset of a disability with around 172% (s.e. 20%). The *t*-test for significance of the impact of a health shock on the onset of a disability is 8.58, which implies

that a health shock not only has a very substantial effect, but the effect is also strongly significant. Not only has a health shock an instantaneous effect on disability rates, but the effect is also long lasting. More specifically, if this individual does not get any other health shock, then with a health shock at age 24, the disability rate at age 40 is 0.141 (s.e. 0.0022). The disability rate at age 40 equals 0.128 (s.e. 0.0018) if the individual never had a health shock. We can conclude from this that health shocks are important for the onset of disabilities and that there are both short-run and long-run effects.

The occurrences of health shocks are relatively rare events. Until age 40 men experience, on average, about 2.4 health shocks and women only 0.8. Model calculations show that health shocks can only account for about 12% of all disabilities at age 40. Hence, the larger part of long standing disabilities arises from a gradual deterioration in health. As discussed in Section 2, the definition of a health shock plays an important role here. In our definition a health shock should lead to a hospital, outpatient or casualty department visit. Health events that did not directly lead to hospital visits are thus included in the gradual deterioration of health. This strict definition of the health shock was required for the identification of the model (see Section 2), but as we now see, this also comes with a price. The instantaneous effect of a health shock on employment rates is negligible. Consider again a non-disabled 24 year old individual. We compare the situation where this individual does not experience a health shock with the situation where s/he does experience a health shock. To get the instantaneous effect of a health shock on employment we assume that the health shock does not lead to a disability (i.e. at his/her 25th birthday the individual is still non-disabled). In this case the employment rate at age 25 is 0.805 (s.e. 0.0018) with the health shock compared to 0.804 (s.e. 0.0018) without a health shock. This difference is very small and not significant. There are, therefore, two potential ways in which a health shock can affect employment in the long-run. First, the model allows for permanent effects and second the effect can run via the onset of a disability. We find that long-run direct effects of health shocks on employment rates are also extremely small. An individual who receives a health shock at age 24 and who does not become disabled in later years has only a 1.1 percentage point lower employment rate at age 40 (the employment rates at age 40 falls from 0.890 (s.e. 0.0020) to 0.879

¹² The long-run effects increases transition probabilities in each year following the health shock, while the instantaneous effects only have an effect in one year. So even though the estimated coefficients of the instantaneous effects are larger, long-run effects can be more larger when the observation period is sufficiently long.

¹³ In fact, we simulate the model for all individuals and compute averages over all individuals.

Table 5
Simulation experiments.

	Fraction	Employment rate age 40	Disability rate age 40	Number of accidents	Effect accident on disability	Effect disability on employment	Association disability and employment
Full population	1.00	0.862	0.124	1.58	172%	−0.144 (0.048)	−0.229
Males	0.50	0.932	0.127	2.40	177%	−0.226 (0.041)	−0.234
Females	0.50	0.790	0.125	0.77	157%	−0.116 (0.124)	−0.253
Low educated	0.58	0.827	0.147	1.57	185%	−0.205 (0.045)	−0.272
High educated	0.42	0.916	0.094	1.60	169%	−0.087 (0.078)	−0.108

Note: The employment rate at age 40, the disability rate at age 40 and the number of health shocks (until age 40) are based on simulating the model. The effect of a health shock on the onset of a disability reflects the effect of an accident at age 24 on the disability rate at the 25th birthday. The effect of a disability on employment is the causal effect of the onset of a disability at age 25 on the employment rate at age 40 computed using the Wald estimator. And the association of disability and employment is based on simulating the model and taking the difference in employment rate at age 40 of those who were and were not disabled at age 25 (this is thus non-causal).

(s.e. 0.0024)). Now suppose that this health shock actually leads to a disability at age 25, then the employment rate at age 40 is 9 percentage point lower (the employment rate falls from 0.890 (s.e. 0.0020) to 0.799 (s.e. 0.0078)).

The results above show that (i) health shocks do not have a direct effect on employment if the individual's disability status remains unchanged and (ii), health shocks have an effect on disability status and thus cause some variation in the disability status. We exploit this additional variation in disability due to the health shock to calculate the causal effect of the onset of a disability on employment. More specifically, we use the model estimates to calculate for each individual in the sample a Local Average Treatment Effect (LATE). Details on the simulation exercise can be found in Appendix C. The treatment that we consider is a health shock at age 24 and no additional health shocks at later ages. The causal effect of the onset of a disability at age 25 on the employment rate at age 40 is then defined as:

$$\frac{\Pr(S_1(40) = 1 | A(24) = 1, \bar{A}(24) = 1) - \Pr(S_1(40) = 1 | A(24) = 0, \bar{A}(24) = 0)}{\Pr(S_h(25) = 1 | A(24) = 1, \bar{A}(24) = 1) - \Pr(S_h(25) = 1 | A(24) = 0, \bar{A}(24) = 0)}$$

$$= -0.144$$

(s.e. 0.048).

The onset of a disability at age 25 causes a significant reduction in the probability of being employed at age 40 by 14.4 percentage points. This estimator exploits the estimated causal effects of the health shocks on the different transition probabilities. However, it provides an estimator for individuals at the margin where a health shock triggers the onset of a disability. The LATE interpretation of the estimator strongly depends on the monotonicity assumption (see Imbens and Angrist, 1994), which can be easily checked in our model.¹⁴ Indeed, we do not find evidence for violation of the monotonicity assumption. Given our model specification violation of monotonicity would be unlikely (but not impossible) given we only model permanent disabilities and that we impose that health shocks are unanticipated.

We can compare this estimate to an alternative estimate, which is derived from taking the difference in employment rates between those who become disabled at age 25 and who do not. The reduction in employment rates at age 40 due to the onset of a disability at age 25 is 0.205 (0.00478).¹⁵ This estimate is substantially larger and much more precise than the estimate for the local average treatment effect above.

4.3. Heterogeneous effects

Recall that the raw data do not show any difference in disability rates between males and females, but males have much higher health shock

¹⁴ For each individual, we can simulate for a health shock at each age if the occurrence of a shock causes that the individual is not less likely to become disabled. This is indeed the case.

¹⁵ So we use $\Pr(S_1(40) = 1 | S_h(25) = 1) - \Pr(S_1(40) = 1 | S_h(25) = 0)$.

risks. Also the types of health shocks differ between males and females. This might imply that health shocks have different effects for males and females. Therefore, we estimate separate models for males and females. We also estimate separate models for high educated (A-level and Degree equivalent) and low educated (O-level and below). High educated and low educated individuals have different types of jobs and may work in different sectors and therefore the causal effects of the onset of a disability on employment may differ. We do not show the parameter estimates for these models, but only show the results from simulation experiments in Table 5.

From Table 5 we see that there is no difference in disability rates at age 40 between men and women (column three), but that women have much lower employment rates (column two). Until age 40 men experience, on average, three times as many health shocks as women (column four) and for men a health shock is more likely to cause a disability (column five). This implies that for men a much larger share of the disabilities is explained by health shocks. Low educated individuals have lower employment rates and higher disability rates than high educated individuals. There is no substantial difference in the health shock rate of high and low educated workers, but for low educated a health shock is more likely to trigger the onset of a disability.

Table 5 shows that the causal effect of a disability on employment differs with respect to gender and education (column six). The onset of a disability has a large and significant causal effect on employment rates of male and lower educated workers. For females and higher educated workers smaller and insignificant effects are found. Note that women have low health shocks rates, so for them we observe much fewer health shocks than for men causing that standard errors are higher. The relatively small and insignificant effect for females may be explained from differences in the type of jobs that men and women are holding. In our analyses we do not make a distinction between part-time and full-time work. Females are more often employed in part-time jobs and it may be easier to continue working in these jobs after the onset of a disability. Also, females have lower participation rates and the marginal working female may be different from her male counterpart. The difference between low and high educated individuals may (again) be due to the different type of jobs and characteristics of the working environment. For instance, continuing working with a chronic condition may be easier for a high educated worker in the service sector than a low educated construction worker.

Column seven of Table 5 reports the difference in employment rates between different groups of workers. This difference is derived from a direct comparison of the employment rate between disabled and non-disabled workers and thus consists of a causal impact of the disability and differences in background characteristics.¹⁶ So this is employment gap merely reflects an association between disability and employment outcomes. For the full sample 14.4 percentage points of the 22.9

¹⁶ More precise, we simulate our model for all workers in our sample and we compute the employment gap at age 40. This association is thus the consequence of both the causal effect of a disability and differences in observed and unobserved background characteristics.

percentage point difference in employment rates between disabled and non-disabled workers is due to a causal effect. The remainder is due to different observed and unobserved background characteristics of disabled and able bodied workers. For males virtually all of the employment gap between non-disabled and disabled workers can be explained by the causal effect, whereas for females the difference in employment rate is primarily due to differences in background characteristics. This implies that for males policies directly aimed at disabled workers, such as the UK DDA or the US ADA, may be more effective than for females. For females other factors are largely explaining the association and hence policies aimed at disabled workers (like workplace accommodations) may not be effective in increasing employment rates. We find similar results for the different education groups. For lower educated workers, the employment difference between non-disabled and disabled workers is substantial (about 27 percentage points) and the larger part of this employment difference can be explained by the causal effect of disability on work. This finding corresponds to Case and Deaton (2005), who argue that health related absences from work are important for the differences in socioeconomic outcomes (like income and work) and that this effect is stronger for workers in manual occupations. Indeed, we find that for higher educated workers the employment gap is much smaller (only about 11 percentage points). The larger part of this employment gap can be explained by the causal effect, but it has to be noted that the causal effect is not significantly different from zero.

5. Conclusions

In this paper we focus on the relation between disability and employment outcomes and particularly the causal effect of disability on employment. We have developed an event-history model that describes transitions between disability and work states and we allowed the transitions rates to be affected by accidents. These health shocks are endogenous in the sense that their arrival depends on the same

observed and unobserved characteristics that affect the transition rates between health and employment states. As measure of the health shocks we have used accidents that caused admittance to a hospital, hospital outpatient or casualty department. We have exploited the unanticipated nature of these accidents to identify the causal effect of such shocks on the onset of a disability and subsequently the causal effect of the onset of a disability on later employment.

The empirical results show that health shocks substantially increase the probability of the onset of a disability, but health shocks have no direct effect on employment (if the health status remains unchanged). Even though a health shock has a large effect on the onset of a disability, the fact that our health shocks are rare events causes that only about 12% of all disabilities at age 40 can be explained from the occurrences of these health shocks. The causal effect of the onset of a disability at age 25 on employment rates at age 40 is in the complete population – 0.144. When estimating separate models for men and women and high and low educated individuals, we found that the onset of a disability has a much larger causal impact on the employment rates of men and low educated than their counterparts.

Finally we have compared the causal effect of the onset of a disability on employment to the association between disability and employment. The latter is the gap in employment between non-disabled and disabled workers, which includes not only the causal effect of disability on employment but also differences in observed and unobserved individual characteristics. Our analyses show that within the complete sample about two-third of the association between disabilities and employment can be explained from a causal effect of disability on employment. This percentage is much lower for women implying that for women observed and unobserved heterogeneity is the primary cause for the association between disability and employment. For males and lower educated workers the larger part of the employment gap can be explained by the causal effect of the disability on employment. This may imply that for these groups policies that aim directly at getting disabled workers back to work can be effective.

Appendix A. Specification, estimation and identification of the model

A.1. Specification

We use a logit specification to model the probability of an accident at age t :

$$q(S_l(t), S_h(t), x(t), v_a) = \frac{\Pr(A(t) = 1 | S_l(t) = s_l(t), S_h(t) = s_h(t), x(t), v_a)}{1 + \exp(x(t)\gamma + \delta_l S_l(t) + \delta_h S_h(t) + v_a)} \tag{1}$$

The parameters δ_l and δ_h describe the effects of being employed and being disabled on the incidence rate of health shocks, respectively. The vector of observed individual characteristics $x(t)$ includes an intercept, a set of socioeconomics characteristics and a fourth order polynomial in age. Recall that because all individuals in the data are born in the same week, the polynomial in age also includes calendar time effects such as business cycle fluctuations.

We define the indicator function $\bar{A}(t)$, which takes on the value 1 if the individual experienced a health shock prior to age t and zero otherwise.¹⁷ The transition probabilities between the different disability and work states are given by:

$$p_{(i,j),(k,m)}(a(t), \bar{a}(t), x(t), v_{(i,j),(k,m)}) = \Pr(S_l(t+1) = i, S_h(t+1) = j | S_l(t) = k, S_h(t) = m, A(t) = a(t), \bar{A}(t) = \bar{a}(t), x(t), v_{(i,j),(k,m)})$$

which we parameterize again as a logit function

$$p_{(i,j),(k,m)}(a(t), \bar{a}(t), x(t), v_{(i,j),(k,m)}) = \frac{\exp(x_t \beta_{(i,j),(k,m)} + \eta_{(i,j),(k,m)} a(t) + \mu_{(i,j),(k,m)} \bar{a}(t) + v_{(i,j),(k,m)})}{1 + \sum_{(i',j') \neq (k,m)} \exp(x_t \beta_{(i',j')(k,m)} + \eta_{(i',j')(k,m)} a(t) + \mu_{(i',j')(k,m)} \bar{a}(t) + v_{(i',j')(k,m)})} \tag{2a}$$

¹⁷ In our model this indicator remains 1 if the individual has more health shocks later.

if $(i,j) \neq (k,m)$ and

$$p_{(k,m),(k,m)}(a(t), \bar{a}(t), x(t), v_{(k,m),(k,m)}) = \frac{1}{1 + \sum_{(i,j) \neq (k,m)} \exp(x_t \beta_{(i,j),(k,m)} + \eta_{(i,j),(k,m)} a(t) + \mu_{(i,j),(k,m)} \bar{a}(t) + v_{(i,j),(k,m)})} \tag{2b}$$

Since disability is an absorbing state this transition probability equals 0 if m is disabled and j is non-disabled.

The parameters $\eta_{(i,j),(k,m)}$ describe the instantaneous effects of a health shock on the different transition probabilities, while the parameters $\mu_{(i,j),(k,m)}$ describe the permanent effects of ever having experienced a health shock. The impact of a health shock can thus be different for individuals in different employment and disability states. Health shocks are not only related to the transition probabilities through these effects, but we also allow for interdependence through the time-invariant unobserved heterogeneity components. The unobserved component in the health-shock probability v_a may be related to the unobserved components in the transitions probabilities $v_{(i,j),(k,m)}$, $\forall i,j,k,m$. We use a random effects specification for the unobserved heterogeneity, and in particular a so-called factor-loading specification to allow for correlations between the different unobserved heterogeneity terms. When estimating the model we take two random factors (w_1, w_2) , which both have two discrete mass points at 0 and 1. The parameters θ_1 and θ_2 denote the probability that the variables w_1 and w_2 take the value 1, respectively. The unobserved heterogeneity terms follow

$$v_a = \alpha_{a,1} w_1 + \alpha_{a,2} w_2 \tag{3a}$$

and

$$v_{(i,j),(k,m)} = \alpha_{(i,j),(k,m),1} w_1 + \alpha_{(i,j),(k,m),2} w_2. \tag{3b}$$

The factor-loading specification implies that the term v_a equals 0 with probability $(1 - \theta_1)(1 - \theta_2)$, $v_a = \alpha_{a1}$ with probability $\theta_1(1 - \theta_2)$, $v_a = \alpha_{a2}$ with probability $(1 - \theta_1)\theta_2$ and $v_a = \alpha_{a1} + \alpha_{a2}$ with probability $\theta_1\theta_2$. The terms $v_{(i,j),(k,m)}$ are defined in a similar way. The unobserved heterogeneity could, for example, pick up the difference between white and blue-collar workers, where the blue-collar workers might have higher probabilities of being disabled and non-employed and a higher incidence rate of health shocks. The parameters α_a and $\alpha_{(i,j),(k,m)}$ determine the sign and the magnitude of the correlation between the accident rate and the transition probabilities. These parameters are estimated along with the other model parameters (as well as the parameters ϑ_1 and ϑ_2).

A.2. Estimation

Since the model is fully parameterized, we use maximum likelihood to estimate the parameters. For each individual, we observe the sequence of realizations $(s_i(T_0 + 1), s_h(T_0 + 1), a(T_0 + 1)), \dots, (s_i(T_1), s_h(T_1), a(T_1))$. In the estimation we condition on the initial labor market status $s_i(T_0 + 1)$ and health status $s_h(T_0 + 1)$ of the individual (i.e. the moment the individual leaves full-time education). The contribution to the likelihood function of for this individual is given by:

$$\begin{aligned} \mathcal{L} = & \sum_{w_1, w_2=0}^1 \theta_1^{w_1} (1-\theta_1)^{1-w_1} \theta_2^{w_2} (1-\theta_2)^{1-w_2} \prod_{t=T_0+1}^{T_1-1} q(s_i(t), s_h(t), x(t), \alpha_{a,1} w_1 + \alpha_{a,2} w_2)^{a(t)} \\ & (1 - q(s_i(t), s_h(t), x(t), \alpha_{a,1} w_1 + \alpha_{a,2} w_2))^{1-a(t)} \\ & P_{(s_i(t), s_h(t)), (s_i(t+1), s_h(t+1))} (a(t), \bar{a}(t), x(t), \alpha_{(s_i(t), s_h(t)), (s_i(t+1), s_h(t+1)), 1} w_1 + \alpha_{(s_i(t), s_h(t)), (s_i(t+1), s_h(t+1)), 2} w_2) \end{aligned}$$

The total likelihood function is the sum of the logarithm of these individual likelihood contributions.

A.3. Identification

Our methodology has been applied when estimating the causal effect of active labor market policies on the reemployment of unemployed workers. [Bonnal et al. \(1997\)](#) estimate the causal effect of training programs for unemployed workers and [Van den Berg et al. \(2004\)](#) evaluate imposing punitive temporary benefits reductions on welfare recipients. We refer to [Abbring and Van den Berg \(2003\)](#) for more formal identification results.

The intuition behind the identification comes from splitting the data in two parts, (i) transition before individuals are hit by a health shock and (ii) transitions after individuals have been hit by a health shock. The assumption of no anticipation implies that we can identify all model parameters except η and μ from data until the moment individuals are hit by the health shock, so part (i) of the data. Part (ii) of the data should thus be used to identify the parameters η and μ . Indeed identification of η and μ comes from comparing transition probabilities before and after experiencing a health shock conditional on observables and unobservables, which have been identified by part (i) of the data. The probability that a health shock occurs can thus vary over time and can differ between individuals based on both observed and time-invariant unobserved characteristics. Without assuming unobserved characteristics to be time-invariant, it would not be possible to identify the effects of health shocks from changes in unobserved characteristics.

Appendix B. Definition of disability

We base our definition of disability on [Currie and Madrian \(1999\)](#) as the mental and physical characteristics that, either constrain normal daily activities, or cause a substantial reduction in productivity

on the job. The NCDS data contains a set of question on health status. Individuals are asked at ages 23, 33 and 42 whether they have a longstanding illness, disability or infirmity which limits their activities compared to people their own age. They are subsequently requested to document whether it limits their daily activities or the

work they can do, the age of the disability onset and the type of disability. Disability types are coded according to the international classification of disease (ICD-9) produced by the World Health Organization (1977).

The ICD is extensively used in health studies and is grouped into 17 broad categories:

1. Infections and parasitic diseases (e.g. tuberculosis, shingles, herpes simplex, glandular fever),
2. neoplasms (e.g. Hodgkin's disease, leukemia),
3. endocrine, nutritional and metabolic diseases and immunity disorders (e.g. obesity, diabetes),
4. diseases of the blood and blood-forming organs (e.g. anemia, coagulation defects),
5. mental disorders (e.g. depression, neurotic disorders, mental retardation),
6. diseases of the nervous system and sense organs (e.g. epilepsy, migraine, blindness, deafness),
7. diseases of the circulatory system (e.g. hypertension, pericarditis, aortic aneurysm),
8. diseases of the respiratory system (e.g. bronchitis, asthma, pleurisy),
9. diseases of the digestive system (e.g. duodenal ulcer, appendicitis, cirrhosis of the liver),
10. diseases of the genitourinary system (e.g. renal failure, cystitis, infertility),
11. complications of pregnancy, childbirth and the puerperium (e.g. spontaneous abortion, ectopic pregnancy),
12. diseases of the skin and subcutaneous tissue (e.g. eczema, psoriasis),
13. diseases of the musculoskeletal system and connective tissue (e.g. rheumatoid arthritis, derangement of joint)
14. congenital anomalies,
15. certain conditions originating in the Perinatal period,
16. symptoms, signs and ill-defined conditions,
17. Injury and poisoning (e.g. fractures, sprains, dislocations, traumatic amputation).

Appendix C. Details of the simulation exercises

The goal of the simulation exercise is to predict within the full population the fraction of individuals that is at a given age in a particular employment or health state, say $S_l(t) = 1$ and $S_h(t) = 0$. In many of the simulations we are interested in health shock that occurs at a moment $\tau < t$, so $A(\tau) = 1$. Our sample consists of n individuals with observed characteristics $x_i = (x_i(T_0), \dots, x_i(T_1))$ and the unobserved heterogeneity has four point of support. This implies

$$\Pr(S_l(t) = 1, S_h(t) = 1 | A(\tau) = 1) = \sum_{w_1, w_2=0}^1 \theta_1^{w_1} (1-\theta_1)^{1-w_1} \theta_2^{w_2} (1-\theta_2)^{1-w_2} \times \left(\sum_{i=1}^n \Pr(S_l(t) = 1, S_h(t) = 1 | A(\tau) = 1, x_i, \alpha_{a,1} w_1 + \alpha_{a,2} w_2) \right).$$

To obtain for each individual the probabilities $\Pr(S_l(t) = 1, S_h(t) = 1 | A(\tau) = 1, x_i, \alpha_{a,1} w_1 + \alpha_{a,2} w_2)$ we exploit that there are only four states. So, for each individual we compute the logit probability of the initial state $S_l(T_0)$ and $S_h(T_0)$ and from there we can use the transition probabilities to exactly compute the probability to be in each of the eight possible states at any moment. When computing these probabilities we take the imposed history of health shocks into account.

Even though we refer to this as simulation exercise, we do not actually simulate, but simply multiply probabilities. The Delta method is used to compute standard errors.

References

- Abbring, J.H., van den Berg, G.J., 2003. The nonparametric identification of treatment effects in duration models. *Econometrica* 71, 1491–1517.
- Acemoglu, D., Angrist, J., 2001. Consequences of employment protection? The case of the Americans with Disabilities Act. *J. Polit. Econ.* 109, 915–957.
- Adams, P., Hurd, M.D., McFadden, D., Merrill, A., Ribeiro, T., 2003. Healthy, wealthy, and wise? Tests for direct causal paths between health and socioeconomic status. *J. Econ.* 112, 3–56.
- Ades, T., 1983. Comparing NCDS4 to the 1981 UK census. National Children's Bureau, National Child Development Studies Working Paper, Fourth Follow-up: No. 11.
- Almond, D., Currie, J., 2011. Killing me softly: the fetal origins hypothesis. *J. Econ. Perspect.* 25 (3), 153–172.
- Bajekal, M., Harries, T., Breman, R., Woodfield, K., 2004. Review of disability estimates and definitions. Department for Work and Pensions, IAD Social Research Division, In-house Reports 128.
- Banks, J., Blundell, R., Bozio, A., Emmerson, C., 2012. "Disability, Health and Retirement in the United Kingdom," NBER Chapters. Social Security Programs and Retirement around the World: Historical Trends in Mortality and Health, Employment, and Disability Insurance Participation. National Bureau of Economic Research, pp. 41–77.
- Bell, D., Heitmueller, A., 2005. The disability discrimination act in the UK: helping or hindering employment amongst the disabled? IZA Working Paper 1476
- Berthoulet, R., 2006. The employment rates of disabled people. Department for Work and Pensions, Research Report No.298.
- Blair, P.S., Fleming, P.J., Bensley, D., Smith, I., Bacon, C., Taylor, E., Gloding, J., Tripp, J., 1996. Smoking and the sudden infant death syndrome: results from 1993–5 case-control study for confidential inquiry into stillbirths and deaths in infancy. *Br. Med. J.* 313, 195–198.
- Bonnal, L., Fougère, D., Sérandon, A., 1997. Evaluating the impact of French employment policies on individual labour market histories. *Rev. Econ. Stud.* 64, 683–713.
- Case, A., Deaton, A., 2005. Broken down by work and sex: how our health declines. In: Wise, D.A. (Ed.), *Analyses in the Economics of Aging*. Chicago University Press.
- Case, A., Fertig, A., Paxson, C., 2005. The lasting impact of childhood health and circumstance. *J. Health Econ.* 24, 365–389.
- Conter, V., Cortinovis, I., Rogari, P., Riva, L., 1995. Weight growth in infants born to mothers who smoked during pregnancy. *Br. Med. J.* 310, 768–771.
- Cropper, M.L., 1977. Health, investment in health, and occupational choice. *J. Polit. Econ.* 85, 1273–1294.
- Currie, J., Hyson, R., 1999. Is the impact of health shocks cushioned by socioeconomic status? The case of low birth weight. *Am. Econ. Rev. Pap. Proc.* 89, 245–250.
- Currie, J., Madrian, B.C., 1999. Health insurance and the labour market. In: Ashenfelter, O., Card, D. (Eds.), *Handbook of Labor Economics 3C*. Amsterdam Elsevier North Holland.
- Currie, J., Thomas, D., 1999. Early test scores, socioeconomic status and future outcomes. NBER Working Papers 6943.
- Deleire, T., 2000. The wage and employment effects of the Americans with Disability Act. *J. Hum. Resour.* 35, 693–715.
- Dupre, D., Karjalainen, A., 2002. Employment of Disabled People in Europe. Eurostat.
- Ehrlich, I., Chuma, H., 1990. A model for the demand for longevity and the value of life extension. *J. Polit. Econ.* 98, 761–782.
- García Gómez, P., López Nicolás, Á., 2006. Health shocks, employment and income in the Spanish labour market. *Health Econ.* 15, 997–1009.
- García-Gomez, P., van Kippersluis, H., O'Donnell, O., van Doorslaer, E., 2013. Long-term and spillover effects of health shocks on employment and income. *J. Hum. Resour.* 48, 873–909.
- Grossman, M., 1972. On the concept of health capital and the demand for health. *J. Polit. Econ.* 80, 223–255.
- Heckman, J.J., Navarro, S., 2007. Dynamic discrete choice and dynamic treatment effects. *J. Econ.* 136, 341–396.
- Hotchkiss, J., 2003. The Labor Market Experience of Workers with Disabilities: the ADA and Beyond. W.E. Upjohn Institute for Employment Research, Kalamazoo.
- Imbens, G.W., Angrist, J.D., 1994. Identification and estimation of local average treatment effects. *Econometrica* 62 (2), 467–475.
- Kapteyn, A., Smith, J.P., van Soest, A., 2007. Vignettes and self-reports of work disability in the United States and the Netherlands. *Am. Econ. Rev.* 97, 461–473.
- Lechner, M., Vasquez-Alvarez, R., 2011. The effect of disability on labour market outcomes in Germany. *Appl. Econ.* 43 (4), 389–412.
- Lindahl, M., 2005. Estimating the effect of income on health and mortality using lottery prizes as exogenous source of variation. *J. Hum. Resour.* 40, 144–168.
- Miguel, E., Kremer, M., 2004. Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica* 72, 159–217.
- Møller-Dano, A., 2005. Road injuries and long-run effects on income and employment. *Health Econ.* 14, 955–970.
- Naeye, R.L., Peters, E.C., 1984. Mental development of children whose mothers smoked during pregnancy. *Obstet. Gynecol.* 64, 601–607.
- NCDS User Support Group, 1991. NCDS5 Report. City University, London.
- Sickles, R.C., Yazbeck, A., 1998. On the demand for leisure and the production of health. *J. Bus. Econ. Stat.* 16, 187–197.
- Smith, J.P., 1998. Socioeconomic status and health. *Am. Econ. Rev. Pap. Proc.* 88, 192–196.
- Smith, J.P., 1999. Healthy bodies and thick wallets: the dual relation between health and economic status. *J. Econ. Perspect.* 13, 145–166.
- Smith, J.P., 2003. Consequences and predictors of new health events. NBER Working Paper 10063.

- Snyder, S.E., Evans, W.N., 2006. The impact of income on mortality: evidence from the social security notch. *Rev. Econ. Stat.* 88 (3), 482–495.
- Van den Berg, G.J., van der Klaauw, B., van Ours, J.C., 2004. Punitive sanctions and the transition rate from welfare to work. *J. Labor Econ.* 22, 211–241.
- Von Gaudecker, H.-M., Garcia Gomez, P., Lindeboom, M., 2011. Health, disability and work: patterns for the working age population. *Int. Tax Public Financ.* 18, 146–165.
- Williams, G.M., O'Callaghan, M., Najman, J.M., Bor, W., Andersen, M.J., Richards, D., Chinlyn, U., 1998. Maternal cigarette smoking and child psychiatric morbidity: a longitudinal study. *Pediatrics* 102, e11.
- World Health Organization, 1977. *International Classification of Diseases, Ninth Revision*. WHO Publications, Geneva.
- World Health Organization, 2004. *Low Birth Weight*. WHO Publications, Geneva.