A Multiple State Model for the Working-age Disabled Population Using Cross-sectional Data

Poontavika Naka[a], Carmen Boado-Penas[b] and Gauthier Lanot[c]

[a] Department of Statistics, Chulalongkorn Business School, Chulalongkorn University, email: poontavika@cbs.chula.ac.th
[b] Department of Mathematical Sciences, Institute for Financial and Actuarial Mathematics (IFAM), University of Liverpool, email: carmen.boado@liverpool.ac.uk
[c] Department of Economics, Umeå University, E-mail: gauthier.lanot@umu.se

Abstract

A multiple state model describes the transitions of the disability risk among the states of active, inactive and dead. Ideally, estimations of transition probabilities and transition intensities rely on longitudinal data; however, most of the national surveys of disability are based on cross-sectional data measuring the disabled status of an individual at one point in time. This paper aims to propose a generic method of the estimation of the expected transition probabilities when the model allows recovery from disability using the UK cross-sectional data. The disability prevalence rates are modelled by taking into consideration the effect of age and time. Under some plausible assumptions concerning the death rates among inactive and active people, the estimated prevalence rates of disability are used to decompose survival probabilities in each state.

Keywords: disability, cross-sectional data, multiple state model, transition probabilities, working-age people

JEL classifications : C35, H53, H55, J19

1. Introduction

Public social spending, which comprises 21% of GDP in 2016 on average across the OECD, is mostly spent on cash benefits related to old age and survivor pensions, incapacity benefits, unemployment, family cash benefits and other social benefits (OECD, 2017). On average, cash income support for the working-age population amounted to 4.4% of GDP in 2013, comprising 1.8% for disability benefits, 1.3% for family cash benefits, 1% for unemployment benefits and 0.3% for other social cash support (OECD, 2014). Notably, for the working-age population, the fiscal cost of incapacity benefit or disability insurance—defined as a periodic income, usually weekly or monthly, paid to an individual who is unable to work due to illness or disablement—has been increasing in several countries, such as Australia, Belgium, France, Iceland, Netherlands, and the United States, due to a substantial growth in disability beneficiary rates (OECD, 2017).

Disability and poor health conditions lead to a decline in labour force participation. Many workers leave the labour market permanently due to health problems or disability, while there are few people with reduced work capacity who remain employed (Jones, 2008; OECD, 2010; Webber and Bjelland, 2015). Very few recipients of disability benefits return to the labour market, even if they have a significant remaining work capacity (OECD, 2009). In the late 2000s, only around 40% of disabled people in OECD countries were employed, and unemployment rates of disabled individuals doubled those of people with no disability (OECD, 2010). At the same time, the high level of unemployment among the disabled population as well as the increasing/larger number of individuals who are receiving long-term sickness and disability benefits raises serious concerns about the sustainability of the public finance of such benefits (Bell and Smith, 2004; McVicar, 2008).

A thorough understanding of the transitions of an individual into and out of a disability state and the accurate estimation of the probability of becoming and remaining disabled are essential data for the government to design the provision of a disability benefit programme, to determine the demand for such programme and to project public expenditure on incapacity benefits.

The logical concept used to describe the transition of disability risk is commonly provided by a multiple state model[1] with relevant states of active (or healthy), disabled (or invalid) and dead (Haberman and Pitacco, 1999; Pitacco, 2014). Although the model typically relies on longitudinal data, most of the national surveys of disability or poor health conditions consist of cross-sectional data measuring the disabled/invalid status of a person at one point in time. Also, the data required for the estimation of transition rates are often missing.

To overcome the limited data for the use of multiple state models, several researchers have recently shown how to derive transition rates across active and disabled states by using disability prevalence rates from cross-sectional data. Rickayzen and Walsh (2002), Leung (2004), Leung (2006) and Hariyanto et al. (2014) identify the functional forms for the one-year deterioration probabilities, i.e. the probabilities of moving to any worse disability level state.

[1] Estimations of transition probabilities and intensities require the total number of transitions from one state to another (e.g. active to disabled, disabled to active, active to dead and disabled to dead), time at transition occurrences and the exposure to risk in each state.

The parameters for each function type are chosen to replicate the observed prevalence rates closely while assuming of a stationary population structure.[2] Nuttall et al. (1994) suggest a multiple state model of health among the elderly considering three states—healthy, disabled and dead—with no transition from the disabled to the healthy state. The disability incidence rates were calculated from the disability prevalence rates and disabled mortality rates. By using the disability prevalence rates, Albarran et al. (2005) compute transition probabilities and survival and death probabilities for the ageing population under the active and disabled states. Also, they employ the annual population mortality rates to decompose the probabilities of death among people with and without disability under some plausible assumptions regarding the relative risk of mortality for each group of individuals.

The abovementioned recent studies have mostly modelled disability rates among the elderly, whereas this paper aims to investigate the evolution among working-age people. We develop a generic estimation method for calculating the transition probabilities in a one-year multiple state model based on disability prevalence rates, hence our method is an extension of Albarran et al.'s (2005) modelling. We apply our method to the UK working-age population using the cross-sectional Labour Force Survey (LFS) to identify employment circumstances and disability prevalence. We then model the disability prevalence and the recovery rates from disability taking into consideration the effect of age and time trends.

Following this introduction, the paper is structured as follows. Section 2 first describes the LFS dataset used to estimate gender- and age-specific disability prevalence and recovery rates and then reports the annual mortality rates for the general population provided by the Human Mortality Database. Section 3 describes our multiple state model and the multiple logistic regression models to estimate disability prevalence rates and the one-year recovery rates. Section 4 the estimated disability rates, one-year recovery rates and transition probabilities are illustrated. Section 5 provides conclusions and additional comments.

[2] The age structure, mortality and birth are constant. The rate of variation between birth and mortality is therefore also constant. Thus, the number of population, of births and of deaths of any age is the fraction of the number of population at an initial age (United Nations, 1968).

2. Data description

The UK LFS is a quarterly survey of the employment circumstances of the UK working-age population, aged 16–59 for women and 16–64 for men.[3] This survey[4] contains self-reported disability data incorporating two definitions of disability: the Disability Discrimination Act[5] (DDA) and the work-limiting disabled. The former applies to any person that currently has a long-term health problem or disability and whose impairment has a substantial and long-term adverse effect[6] on his/her ability to undertake normal day-to-day activities[7]. The latter applies to any work-limiting disabled individual who has a long-term health problem or disability relating specifically to working life and whose impairments affect either the kind or amount of work he/she might do. In the LFS there is one question about the current respondent's disability. The possible answers by the respondent are 1) both DDA (current disability) and work-limiting disabled, 2) DDA disabled (current disability) only, 3) work-limiting disabled only and 4) not disabled.[8]

The LFS surveys any respondent every three months for five consecutive quarters. This allows us to have a one-year observation of transitions among the different states. The LFS provides information on the individual's labour force status, i.e. employed, unemployed or economically inactive. The overall sample size of the cross-sectional LFS dataset over the period 1999–2011 consists of 576,402 people, of which 288,576 are males aged 16–64 and 287,826 are females aged 16–59. In each dataset, we use the given person-weight variable to gross up the survey estimates to population totals. This sampling weight is based on the number of similar people in the whole population in the particular time of the survey and controlling for age and sex. We

[3] Until April 2010, the state pension age in the UK was 60 for women and 65 for men.

[4] There are a few national surveys on disability in the UK. For example, in 1986 the Office of Population Censuses and Surveys (OPCS) classified disabled children and adults according to ten degrees of disability. However, this dataset is out of date and given that it was carried out only in a single year is unable to illustrate trends in disability. The Understanding Society panel, wave 1-6, 2009-2015 is one of longitudinal study covering the questions on self-reported longstanding illness or disability and activity limiting condition; however, at the time we are conducting this research only three consecutive waves have been released.

[5] The Disability Discrimination Act (1995) (DDA), which protects disabled people from discrimination, was repealed and replaced by the Equality Act 2010, except in Northern Ireland where this Act is still applied.

[6] See Equality Act 2010: Guidance on matters to be taken into account in determining questions relating to the definition of disability, Section B: Substantial, p. 14–26.

[7] See Equality Act 2010: Guidance on matters to be taken into account in determining questions relating to the definition of disability, Section D: Normal day-to-day activities, p. 34–47.

[8] Since the answers of the LFS are based on respondents' self-assessment there is no more information on medical tests or the degree of disability. In this paper, as we want to classify the individuals as disabled (inactive) and non-disabled (active), we include all respondents who selects (1), (2) or (3) as disabled while we classify as non-disabled those who respond (4). We acknowledge that a limitation in our paper is the fact that we cannot distinguish the degree of disability of the individuals and we merge all disabled individuals into the same category, i.e. inactive state.

estimate the total number of people in the working-age population and the disabled population as shown in Table 1.[9]

Table 1: Number of inactive working-age population estimates and disability prevalence rates by gender and year 1999–2011

Year	Male			Female			Overall		
	Inactive Population	Total Population[10]	Disability rate (%)	Inactive Population	Total Population	Disability rate (%)	Inactive Population	Total Population	Disability rate (%)
1999	1,776,477	18,323,457	9.6951	1,722,695	17,213,707	10.0077	3,499,172	35,537,165	9.8465
2000	1,772,596	18,422,980	9.6217	1,803,039	17,327,492	10.4057	3,575,634	35,750,472	10.0016
2001	1,797,411	18,556,952	9.6859	1,786,135	17,462,581	10.2284	3,583,546	36,019,533	9.9489
2002	1,880,029	18,680,082	10.0643	1,802,317	17,539,222	10.2759	3,682,345	36,219,304	10.1668
2003	1,778,020	18,818,881	9.4481	1,888,851	17,668,803	10.6903	3,666,871	36,487,684	10.0496
2004	1,789,503	18,946,537	9.4450	1,831,128	17,798,479	10.2881	3,620,631	36,745,016	9.8534
2005	1,800,498	19,145,553	9.4043	1,790,341	17,954,150	9.9717	3,590,838	37,099,703	9.6789
2006	1,812,026	19,339,180	9.3697	1,825,118	18,107,852	10.0792	3,637,145	37,447,032	9.7128
2007	1,849,885	19,532,406	9.4708	1,836,278	18,189,036	10.0955	3,686,163	37,721,442	9.7721
2008	1,870,487	19,704,044	9.4929	1,737,931	18,256,260	9.5196	3,608,418	37,960,304	9.5058
2009	1,842,044	19,814,587	9.2964	1,811,686	18,331,005	9.8832	3,653,730	38,145,592	9.5784
2010	1,951,609	19,910,234	9.8020	1,851,813	18,408,702	10.0594	3,803,422	38,318,937	9.9257
2011	2,088,107	19,955,266	10.4639	1,816,924	18,241,314	9.9605	3,905,032	38,196,581	10.2235

Source: The authors' calculation based on the LFS dataset

In this paper, we link self-assessed disabled people with the labour force status—unemployed or economically inactive—as a proxy for the number of disabled individuals who are entitled to receive incapacity benefits (state 'Inactive' in Figure 1). The remaining individuals, i.e. non-disabled people and employed disabled people, act as a proxy for the number of non-recipients of disability/incapacity benefits (state 'Active' in Figure 1). We use the cross-sectional dataset of each first quarter (January–March) over the period 1999–2011 to model trends in disability

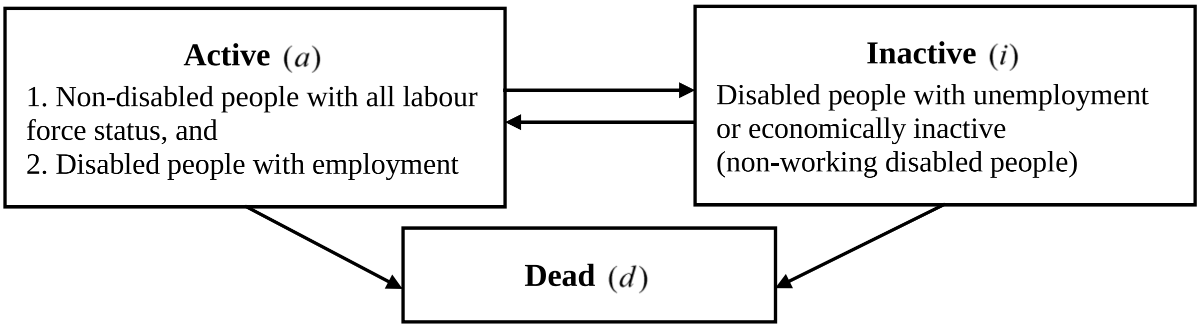
---

[9] The estimator for the number of individuals in the population is the sum of person-weight provided in the LFS dataset. It can be expressed as $\hat{Y} = \sum_{j=1}^{n} w_j$; where w_j is a sampling weight for the *j*-th sampled individual from the population, $j = 1, 2, \ldots, n$ and n is the number of observations in the sample.

[10] The total number of working-age population estimates in each year are approximately equal to the number of population estimates provided by the HMD.

prevalence rates, while we estimate one-year recovery rates for disabled people using the status information drawn from interviews in quarter 1 and 5.[11]

Figure 1: Three-state model of working-age people



In the following subsection, we clarify the characteristics of the datasets, including the disability prevalence rates, recovery rates and mortality rates of the working-age population that are used in this study.

2.1. Disability prevalence rates of the working-age population

The prevalence rate of disability at age x is computed as the total number of disabled individuals aged x divided by the total population of age x. On average, disability prevalence rates have remained quite constant (between 9–10%) over the whole period of analysis and the rates of women are higher than men (see Table 1).

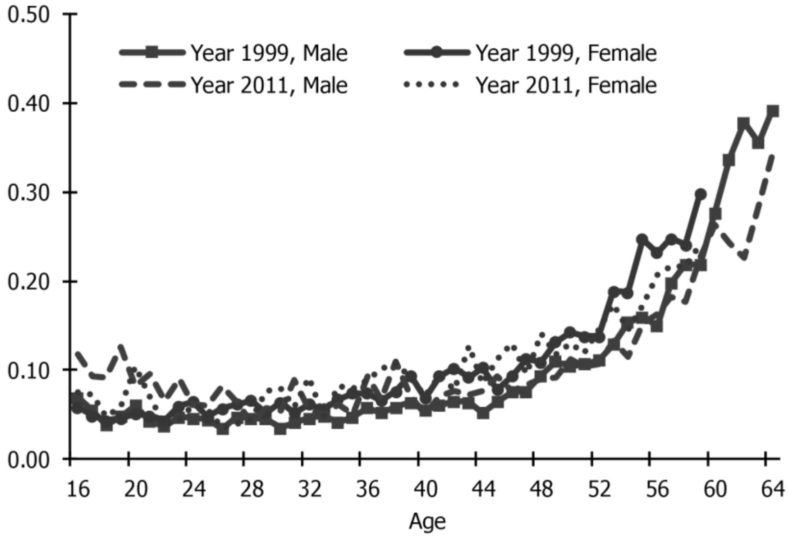
As shown in Figure 2, there is a noticeable age pattern in the disability rates, with lower rates among the young individuals and an increase in older ages. In early adulthood, aged 16–25, the disability prevalence rates for males are slightly larger than for females, whereas the prevalence rates tend to be larger among women during the middle age, i.e. around age 40. As a result, the disability prevalence rates are associated with age and gender. The time effect might also have an influence on changes in rates, although the trends in disability rates are quite unclear over time.

[11] We merge all five-quarter longitudinal datasets over the period 1999–2011 since the number of respondents who are being disabled in each year is relatively small. As a result, the modelling of gender- and age-specific recovery rates ignores time effects and the rates remain constant over time.

2.2. One-year recovery rates of the working-age population

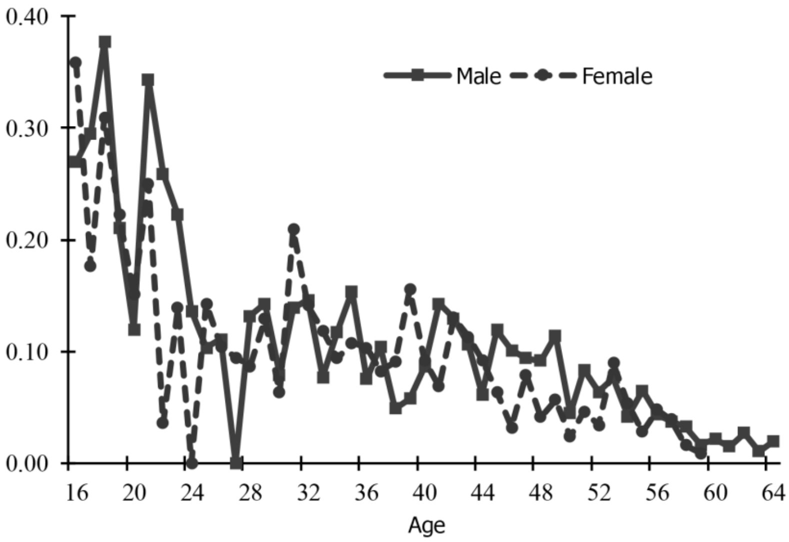
The one-year recovery rate from the disable state at age x is represented by the ratio between the total number of disabled population aged x transferring to the active state over one year and the total number of the disabled population of age x. The recovery rates, as shown in Figure 3, decline with age, from around 4% to 0.5% for males and females; however, it is unclear whether there are gender differences in the recovery rates.

Figure 2: Observed disability prevalence rates by age and gender in 1999 and 2011



Source: The authors' calculation based on the LFS datasets

Figure 3: Observed one-year recovery rates among men and women, 1999–2011, by age

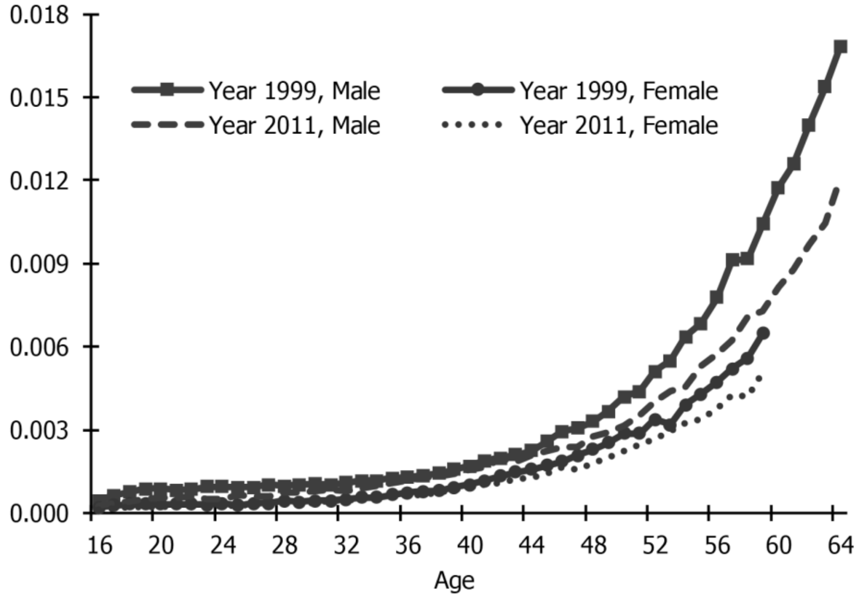


Source: The authors' calculation based on the LFS datasets

2.3. Annual mortality rates

As we only have information of the mortality rates for the general population, this subsection describes the methodology used to decompose the probabilities of death among the active and inactive people under the multiple state model shown in Figure 1. The mortality rates for males, as shown in Figure 4, are consistently above those of females, while the mortality rates have dropped gradually at all ages and for both sexes over the period 1999–2011.[12]

Figure 4: Age-specific mortality rates (q_x) by gender in 1999 and 2011



Source: Human Mortality Database

3. Model specification

This section describes the discrete-time multiple state model to compute for each of the three states, particularly the following two types of probabilities: those associated with remaining in the same state and those related to transitions between states. Our model allows for recovery from the inactive to the active state by extending Albarran et al.'s (2005) approach, which introduced some assumptions about the relative mortality ratio among disabled and non-disabled people to decompose the probability of death in any state.

---

[12] The mortality improvement arises from economic development, progress in health technology, better access to health care services, rising living standards, improved lifestyles and a shift in the leading causes of death and illness from the infectious and parasitic diseases to non-communicable diseases and chronic conditions, especially cancers and diseases of the circulatory system (Howse, 2006; Soubbotina and Sheram, 2000; WHO, 2011).

3.1. The multiple state model for working-age people

In the three-state model of working-age people, as shown in Figure 1, the possible transitions are as follows: (i) disablement, i.e. transition from the 'active' to the 'inactive' state; (ii) recovery, i.e. transition from the 'inactive' to the 'active' state; (iii) death of an active individual, i.e. transition from the 'active' to the 'dead' state and (iv) death of an inactive person, i.e. transition from the 'inactive' to the 'dead' state. The actuarial notations of one-year transition probabilities and the equations used to estimate the transition probabilities and probabilities in any state are included in the next subsection.

3.1.1. The one-year transition probabilities

We apply a discrete time of three states of working-age people model in a one-year period according to Haberman and Pitacco (1999) and Pitacco (2014). We also assume that, except the possible death of an individual, no more than one transition occurs during one particular year. The fundamental relations of one-year transition probabilities related to an active individual and an inactive individual age are explained in the following notations (see more details in Appendix A):

$$p_x^{jj} + p_x^{jk} = p_x^j \quad (1)$$

$$q_x^{jj} + q_x^{jk} = q_x^j \quad (2)$$

$$p_x^j + q_x^j = 1 \quad (3)$$

$$p_x^{jk} + q_x^{jk} = w_x^{jk} \quad (4)$$

$$p_x^{jj} + q_x^{jj} = 1 - w_x^{jk} \quad (5)$$

where p_x^{jk} denotes the probability that a person aged x in a state j is alive in a state k at age $x+1$;

q_x^{jk} denotes the probability that a person aged x in a state j dies within one year in a state k;

p_x^j denotes the probability that a person aged x in a state j is alive at age $x+1$;

q_x^j denotes the probability that a person aged x in a state j dies within one year;

w_x^{jk} denotes the probability that a person aged x in a state j moves to a state k;

j,k represent any state of a 'active' and i 'active', $j \neq k$.

As we assume that there is no more than one transition occurring during one year, apart from the possible death, consequently the p_x^{aa} and p_x^{ii} represent the probabilities of remaining in the active and inactive state, respectively, from age x to $x+1$. Furthermore, the probability of becoming inactive is equivalent to w_x^{ai} and the probability of recovery from an inactive to an active state within one year, w_x^{ia}, is represented by the estimated one-year recovery rate from the logistic regression model in Section 3.2.

3.1.2. Estimating the survival and transition probabilities

In this subsection we explain how to estimate the transition probabilities and the probabilities of remaining in the same state throughout one year, thereby extending Albarran et al.'s (2005) approach. Because of the lack of information of mortality rates across subpopulation, Albarran et al. (2005) disaggregate the mortality rates for the general population at age x, q_x into the mortality rates of disabled and non-disabled people. The assumptions of the hazard ratio[13] of disability on mortality are also supposed to approximate the probabilities and the transition probabilities among the active and the inactive population.

Initially, we decompose the mortality rate for the general population at age x, q_x into the weighted average of the mortality rate for the active people, q_x^a, and the inactive people, q_x^i, with the proportion of active and inactive people, respectively (Majer et al., 2013), as defined in the following expression:

$$q_x = (1-v_x)q_x^a + v_x q_x^i = (1-v_x)(q_x^{aa} + q_x^{ai}) + v_x(q_x^{ii} + q_x^{ia}) \qquad (6)$$

We use the annual mortality rates of the UK working-age population by age and gender, which were obtained from the Human Mortality Database.[14] The proportion of inactive people is measured by the probability of being inactive, v_x, and this is clearly equivalent to the prevalence rate of disability at a particular age x that is estimated by the multiple logit regression model, which is explained in Section 3.2.

[13] The hazard ratio of disability on mortality is equivalent to the relative mortality risk of inactive people versus active people regarding the standard Cox proportional hazard assumption: $q_x^i = HR \cdot q_x^a$, where q_x^i and q_x^a is the mortality rate of inactive people and active people at age x, respectively, while HR is the hazard ratio of disability on mortality.

[14] www.mortality.org

The probability that an individual aged x survives up to age $x+1$, p_x can also be identified as follows:

$$p_x = (1-v_x)(p_x^{aa} + p_x^{ai}) + v_x(p_x^{ii} + p_x^{ia}) \quad (7)$$

We then make three assumptions regarding the hazard ratio of inactive people on mortality, which is the ratio between the mortality rate of inactive and active people. These three common assumptions are defined as follows:

Assumption 1: $q_x^{ai} = k_1 w_x^{ai} q_x^{ii}$; $0 < k_1 \leq 1$

Assumption 2: $q_x^{ia} = k_2 w_x^{ia} q_x^{aa}$; $0 < k_2 \leq 1$

Assumption 3: $q_x^{aa} = k_3 q_x^{ii}$; $0 < k_3 \leq 1$

According to Albarran et al. (2005), we follow their Assumption 1 in term of the age distribution of becoming inactive. For Assumption 2, we establish the ratio between the two death probabilities, q_x^{ia} and q_x^{aa}, as a function of the age distribution of recovery. Because of the work of Albarran et al. (2005) focusing on the elderly, they assume the ratio among two mortality rates q_x^{ii} and q_x^{aa} is a function of age and the gap in both mortality rates tends to increase with age in the old-age group. However, they point out that this may not be true for the whole population. Majer et al. (2013) additionally found that there is no significant age interaction or time trend in the Cox proportional hazard ratios between the mortality rates of the Dutch non-disabled and disabled populations, with the constant ratio of 0.54 and 0.58 for men and women, respectively. As a result, in Assumption 3, we require that the hazard ratios for the mortality risk among active and inactive populations are constant over age.

The mortality rates for the inactive population are generally higher those of active[15] population, which means the ratio among the mortality rates of active and inactive population are lower than 1. Also, the probability that an active (or inactive) individual dies in the different state is

[15] See Majer et al (2011), Forman-Hoffman et al (2015), Bahk et al (2019), amongst others.

likely to be lower than the probability of active (or inactive) dying in the same state[16], i.e. $q_x^{ii} > q_x^{ai}$ and $q_x^{aa} > q_x^{ia}$. Therefore, the range of k_1, k_2, and k_3 is set between 0 and 1.

Next, a stationary population assumption requires that the number of inactive people at age $x+1$ is the sum of the number of active people aged x surviving in the same active state at age $x+1$ and the number of active people aged x surviving in the inactive state at age $x+1$. It is expressed as follows:

$$v_{x+1} p_x = v_x p_x^{ii} + (1-v_x) p_x^{ai} \tag{8}$$

By means of substituting the expression (5) and Assumption 1 in the expression (8), we obtain the relationship:

$$v_{x+1}(1-q_x) = v_x(1-w_x^{ia} - q_x^{ii}) + (1-v_x)(w_x^{ai} - k_1 w_x^{ai} q_x^{ii}) \tag{9}$$

which yields the probability of becoming inactive between age x and $x+1$:

$$w_x^{ai} = \frac{v_{x+1}\left(1-q_x\right) - v_x\left(1-w_x^{ia} - q_x^{ii}\right)}{\left(1-v_x\right)\left(1-k_1 q_x^{ii}\right)} \tag{10}$$

Combining Assumption 1, 2 and 3 and expression (6), we then obtain the probability that an inactive person at age x dies within one year while he/she is still inactive:

$$q_x^{ii} = \frac{q_x}{\left(1-v_x\right)k_3 + \left(1-v_x\right)k_1 w_x^{ai} + v_x + v_x k_2 k_3 w_x^{ia}} \tag{11}$$

Finally, substituting expression (10) in (11), we obtain a quadratic equation of q_x^{ii} as follows:

$$A\left(q_x^{ii}\right)^2 + B q_x^{ii} - q_x = 0 \tag{12}$$

where $A = -k_1 k_3 \left[\left(1-v_x\right) + v_x k_1 k_2 w_x^{ia}\right]$

and $B = k_3\left(1 - v_x + v_x k_2 w_x^{ia}\right) + k_1 v_{x+1}\left(1-q_x\right) + k_1\left(q_x - v_x + v_x w_x^{ia}\right) + v_x$

[16] This is due to the fact that the first probability also requires a transition into a different state.

The equation admits two real positive solutions. However, we choose the unique solution that lies in the (0, 1) interval, as shown in Appendix B.

Replacing the known values of q_x, v_x and w_x^{ia} with different values of k_1, k_2 and k_3 in the solution of equation (12), we obtain the probability that an inactive person dies while he/she is in an inactive state, q_x^{ii}. We then compute the probability of becoming inactive, w_x^{ai}, from the equation (10) and the probability of death in any state q_x^{ai}, q_x^{ia} and q_x^{aa} from the Assumption 1, 2 and 3.

3.2. Multiple logistic regression model

The LFS datasets contain a binary outcome indicator of the disability event occurrence (i.e. non-disabled or disabled status) and of the case of recovery from disabled to non-disabled (i.e. non-recovery or recovery status). The logistic regression is a popular model for binary dependent variables that allows us to estimate the probability of the event of interest (De Jong and Heller, 2008; Frees, 2009; Hosmer et al., 2013; Guillen, 2014).

We employ the logistic regression models to capture the occurrence of disability and recovery events separately for males and females by taking into account the effect of age as a polynomial function (Renshaw and Haberman, 1995; Fong et al., 2015) and of time trends (Renshaw and Haberman, 2000). The estimations of disability prevalence rates and one-year recovery rates are explained in the following subsection.

3.2.1. Estimating the disability prevalence rates, v_x

In order to estimate gender- and age-specific disability prevalence rates of the working-age population in each calendar year over the period 1999–2011, we apply the logistic regression with age and time trend as the predictor variables. The binary outcome of the event that the *n*-th person is being inactive, y_n, is defined as follows:

$$y_n = \begin{cases} 1 & \text{if the n-th person is inactive with probability } v_n \\ 0 & \text{if the n-th person is active with probability } 1\text{-}v_n \end{cases}$$

The logistic regression model of disability prevalence rates is based on a polynomial of age with degree 4 and a time trend, as shown in the next equation below[17]. The model is analysed for males and females separately.

The logistic regression of estimation disability prevalence rates is defined as follows:

$$\text{logit }(v_{n,t}) = \ln\left(\frac{v_{n,t}}{1-v_{n,t}}\right) = \alpha + \beta_1 age + \beta_2 age^2 + \beta_3 age^3 + \beta_4 age^4 + \beta_5 t \qquad (13)$$

where $v_{n,t}$ represents the disability prevalence rate of the *n*-th person in calendar year t, age is the age of the individual and t is the calendar year, i.e. 0,1, …,12 corresponding to the year between 1999–2011.

The fitted gender- and age-specific disability prevalence rate at age x in each calendar year over the period 1999–2011 is expressed as follows:

$$\hat{v}_x = \frac{\exp(\hat{\alpha} + \hat{\beta}_1 age + \hat{\beta}_2 age^2 + \hat{\beta}_3 age^3 + \hat{\beta}_4 age^4 + \hat{\beta}_5 t)}{1 + \exp(\hat{\alpha} + \hat{\beta}_1 age + \hat{\beta}_2 age^2 + \hat{\beta}_3 age^3 + \hat{\beta}_4 age^4 + \hat{\beta}_5 t)} \qquad (14)$$

3.2.2. Estimating the one-year recovery rates, w_x^{ia}

The binary outcome of the event that the *n*-th disabled person recovers to an active state over a one-year period, z_n, is defined as follows:

$$z_n = \begin{cases} 1 & \text{if the n-th inactive person recovers to an active state with probability } w_n^{ia} \\ 0 & \text{if the n-th inactive person is still inactive with probability } 1\text{-}w_n^{ia} \end{cases}$$

The logistic regression model is explained by gender and a quadratic function of age using the following equation[18]:

$$\text{logit }(w_n^{ia}) = \ln\left(\frac{w_n^{ia}}{1-w_n^{ia}}\right) = \alpha + \beta_1 gender + \beta_2 age^2 \qquad (15)$$

[17] Different polynomial forms have been carried out but we only show results for the best fit to the data.

[18] Different polynomial forms have been carried out but the model with gender and degree 2 of age as predictors is the best fit to the data.

where $gender$ is a dummy variable with 1 for males and 2 for females and age is the age of the *n*-th disabled individual.

The constant estimated gender-specific one-year recovery rate at age x over the period 1999–2011 is expressed as follows:

$$\hat{w}_x^{ia} = \frac{\exp(\hat{\alpha} + \hat{\beta}_1 gender + \hat{\beta}_2 age^2)}{1 + \exp(\hat{\alpha} + \hat{\beta}_1 gender + \hat{\beta}_2 age^2)} \tag{16}$$

The logistic regression of both equations (13) and (15) are fitted to the data by using maximum likelihood methods to obtain the estimates of parameters, i.e. the intercept (α) and coefficients $(\boldsymbol{\beta})$. Then, we compute the fitted disability rates from (14) and the fitted one-year recovery rates from (16) by substituting the estimated parameters α and $\boldsymbol{\beta}$, which are illustrated in the following section.

4. Results

In this section we discuss the results for the estimated gender- and age-specific disability prevalence rates and the one-year recovery rates over the period of 1999–2011. All estimated rates are included in the one-year multiple state model to generate the transition probabilities in each state (i.e. active, inactive and dead). The probabilities of death among each state are also computed, based on the assumptions regarding the relative mortality risk between inactive and active people described in Section 3.1.

4.1. The estimated disability prevalence rates, v_x

The results of the estimates for the unknown parameters for men and women are shown in Table 2. In the logistic regression model, parameters are interpreted in terms of logit rather than directly in the response variable. Then, these estimated parameters are calculated following the equation (14) to produce the fitted disability prevalence rates by age and gender in each calendar year.

As shown in Table 2, all parameters are statistically significant at the 90% confidence level. This means that age and time trends have an influence on the probabilities of the UK working-age population being disabled. The prevalence rate of disability rises with age and is higher for women than men. Conversely, the young men aged 18–27 have higher disability prevalence rates than the young women. The trends in disability rates among men and women have dropped slightly over time due to the negative value of the coefficient of t.

In Figure 5, the fitted disability rates of young men aged 16–28 in 1999 decreased from 0.0795 to 0.0483, whereas the rates of men aged 29–64 increased from 0.0486 to 0.3723. For females, the fitted rates were lower, i.e. ranging between 0.0565–0.2808 for the age interval 16–59. The disability rates in 2011 slightly decreased from 1999 for both sexes. The fitted rates have the same trend as the disability living allowance claimant rates, which is the social welfare for disabled people. However, for all age groups the fitted rates from our model produce higher values than the rates of claimants since beneficiaries who are entitled to receive benefits have to meet strict conditions[19]. As a result, the number of recipients is likely to be lower than the number of people with self-reported disability.

Table 2: The logistic regression model of disability prevalence rates

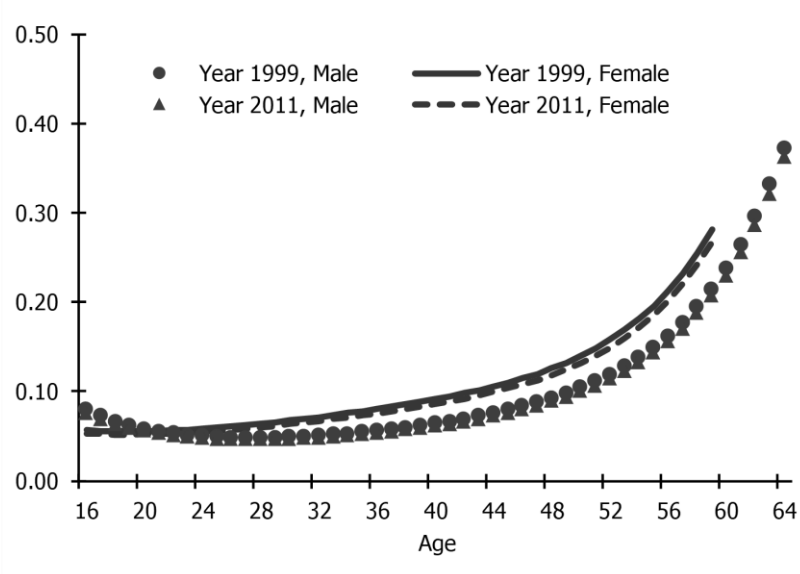
Parameter	Estimate	Std.Error	z value	Pr > |z|
Males				
α (intercept)	1.9380	0.5292	3.6600	0.0000^{***}
β_1 (age)	-0.4964	0.0615	-8.0700	0.0000^{***}
β_2 (age^2)	0.0179	0.0025	7.2000	0.0000^{***}
β_3 ($age^3/1000$)	-0.2790	0.0421	-6.6300	0.0000^{***}
β_4 ($age^4/10000$)	0.0174	0.0025	6.8200	0.0000^{***}
β_5 (t)	-0.0038	0.0019	-2.0400	0.0420^{**}
Chi-square = 13735.58				
Pr > Chi-square = 0.0000				
Females				
α (intercept)	-1.1506	0.6586	-1.7500	0.0810^{*}
β_1 (age)	-0.2339	0.0792	-2.9500	0.0030^{***}
β_2 (age^2)	0.0111	0.0034	3.3000	0.0010^{***}
β_3 (age^3)	-0.0002	0.0001	-3.4800	0.0000^{***}
β_4 ($age^4/10000$)	0.0154	0.0004	3.9400	0.0000^{***}
β_5 (t)	-0.0057	0.0018	-3.1200	0.0020^{**}
Chi-square = 7553.25				
Pr > Chi-square = 0.0000				

Source: The authors' own source using the LFS dataset and the logistic regression eq. (13)

[19] See the rates of claimants receiving Disability Living Allowance (DLA) from the Department for Work & Pensions (DWP) statistics tabulation tool over the period 2002-2011. The values of the disability rates differ due to the different nature of both datasets, i.e. while LFS is a self-assessed dataset, the claimants of DWP are examined by an independent healthcare professional. Also, under the DWP, the claimants must have a long-term health condition or disability and face difficulties with 'daily living' or getting around. These difficulties must be longer than 3 months and are expected to last at least 9 months.

Note: $*p-value<0.10;\quad **p-value<0.05;\quad ***p-value<0.01$

Figure 5: Estimated disability prevalence rates by age and gender in 1999 and 2011



Source: The authors' own source from the logistic regression model in the eq. (14)

4.2. The estimated one-year recovery rates, w_x^{ia}

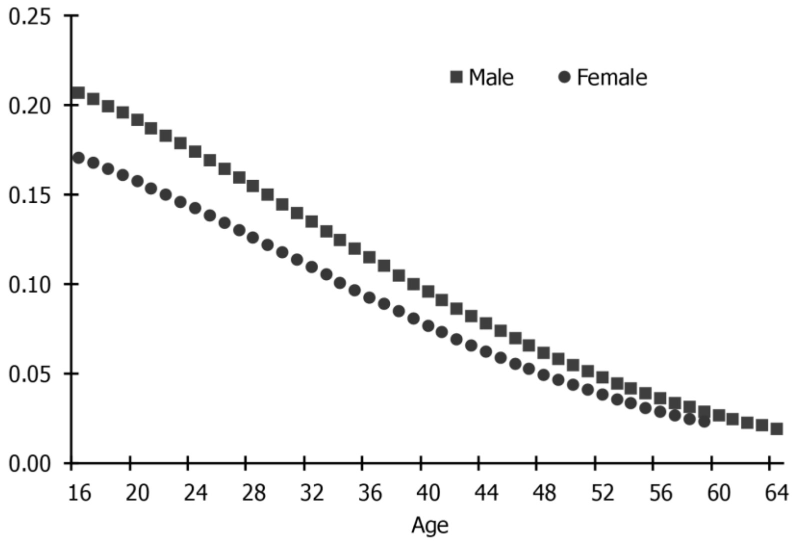
As shown in Table 3, the variables age and gender are statistically significant to model the one-year recovery rates for inactive people. We can see that the one-year recovery rates gradually decrease with the quadratic form of ages. The coefficient of gender variable is negative, which means that more men recover their health and get a job during the course of one year than women. This is not surprising since most women with disabilities encounter barriers in entering the labour market and often experience employment disadvantages, such as inequality in hiring, promotion standards and payment (O'Reilly, 2007). As a result, many women do not desire to return to work.

Table 3: The logistic regression model of one-year recovery rates

Parameter	Estimate	Std.Error	z value	Pr > |z|
α (intercept)	-1.1705	0.1071	-10.9300	0.0000***
β_1 (gender)	-0.2379	0.0959	-2.4800	0.0130**
β_2 (age^2)	-0.0007	0.0000	-15.7100	0.0000***
Chi-square = 247.91				
Pr > Chi-square = 0.0000				

Source: The authors' own source using the LFS dataset and the logistic regression eq. (15)
Note: $*p-value<0.10$; $**p-value<0.05$; $***p-value<0.01$

Figure 6: Estimated one-year recovery rates by age and gender over 1999–2011



Source: The authors' own source from the logistic regression model in the equation (16)

The fitted recovery rates over one year, as shown in Figure 6, decrease from 0.2071 to 0.0194 for men and from 0.1707 to 0.0230 for women. Our fitted rates are consistent with the one-year claim duration recovery rates provided by the Society of Actuaries: 2008 Long Term Disability Experience Study Report. The report gathers and analyses historical industry data on long-term disability claims of the US insurance companies between 1997 and 2006. The analysis shows that, on average, the rates decreased with increasing age, 0.1973 for under the age of 25, and dropped to 0.0306 for the 60–64 age band. The fitted rates from our model for males and females are approximate to these experience rates.

4.3. Estimated annual survival and transition probabilities

Taking into account the one-year multiple state model to estimate the transition probabilities and the probabilities of remaining in the same state, we use the annual gender- and age-specific mortality rates (q_x) for the general population, the estimated disability prevalence rates (v_x) and the estimated one-year recovery rates (w_x^{ia}) from the logistic regression model. We calculate the probabilities of becoming inactive (w_x^{ai}), the probabilities of death in any state (

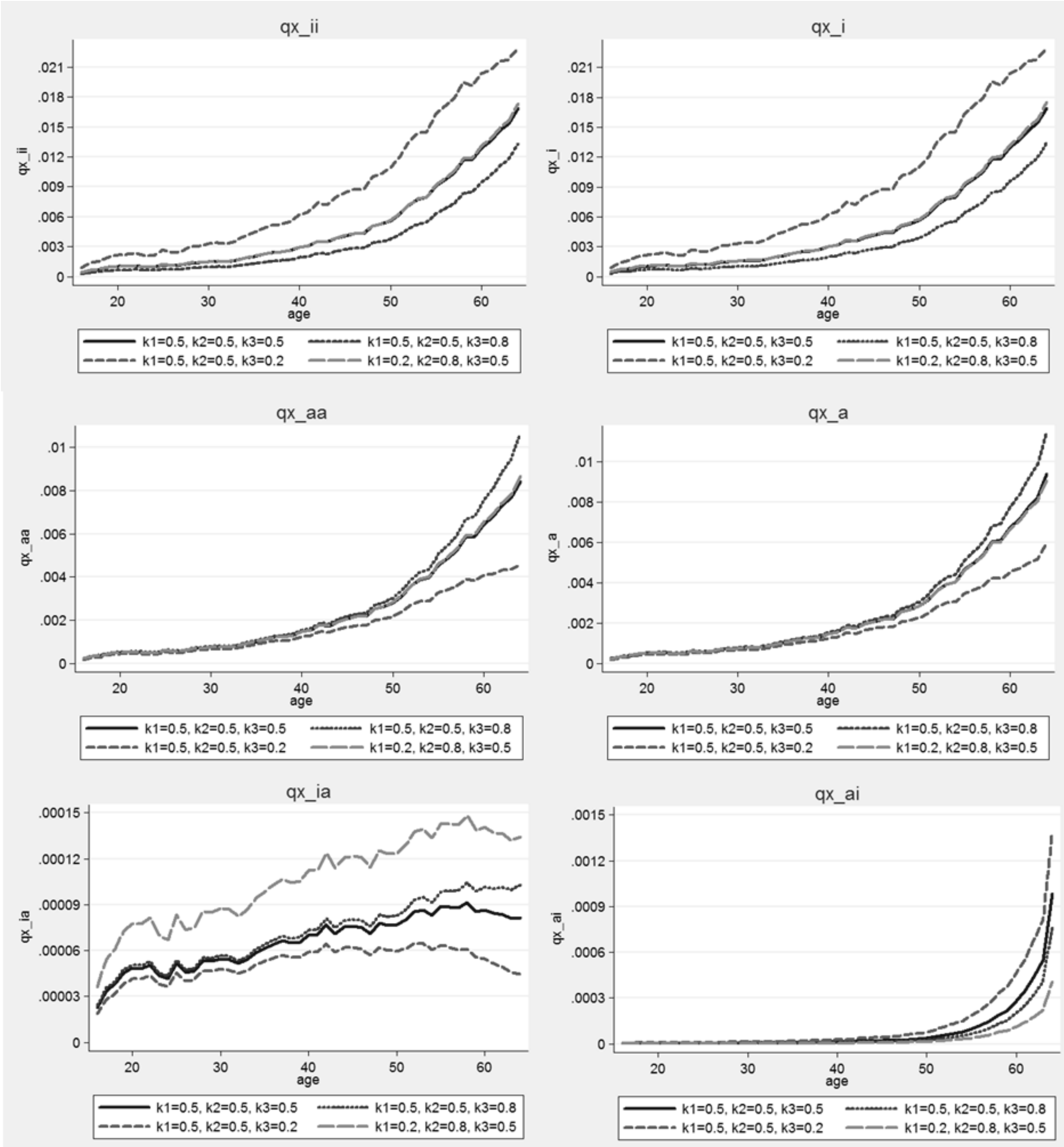
$q_x^{ai}, q_x^{ia}, q_x^{ii}$ and q_x^{aa}) and the probabilities of remaining in the same state (p_x^{ii} and p_x^{aa}). In this paper, we illustrate only one typical year, 2011, to perform the results in accordance with the various values of k_1, k_2 and k_3 from Assumption 1, 2 and 3.

We show the probabilities of death in any state with different values of $k_1 = k_2 = k_3$ and $k_1 \neq k_2 \neq k_3$ in Figure 7. The graphic of the probabilities of inactive people aged x dying in an inactive state (q_x^{ii}) shows that the value of k_3, which is equivalent to the relative mortality risk ratio among active and inactive people, is negatively correlated with the value of q_x^{ii}. The lower value of k_3 produces the higher value of q_x^{ii}; for example, the value of q_x^{ii} under the scenario of the lowest $k_3 = 0.2$ together with $k_1 = 0.5$ and $k_2 = 0.5$ is higher than the other scenarios. Conversely, there is a positive relationship between the value of k_3 and the probability of death of an active person. The higher value of k_3 gives the lower value of q_x^{aa}.

In the case of changes in the values of k_1 and k_2 while keeping the same value of k_3 at 0.5, the values of q_x^{ii} and q_x^{aa} are almost unchanged; for example, the results given $k_1 = 0.5, k_2 = 0.5, k_3 = 0.5$ against $k_1 = 0.2, k_2 = 0.8, k_3 = 0.5$ are very similar. Thus, the k_1 and k_2 have very minor effect on the probabilities of death in the same state q_x^{ii} and q_x^{aa}.

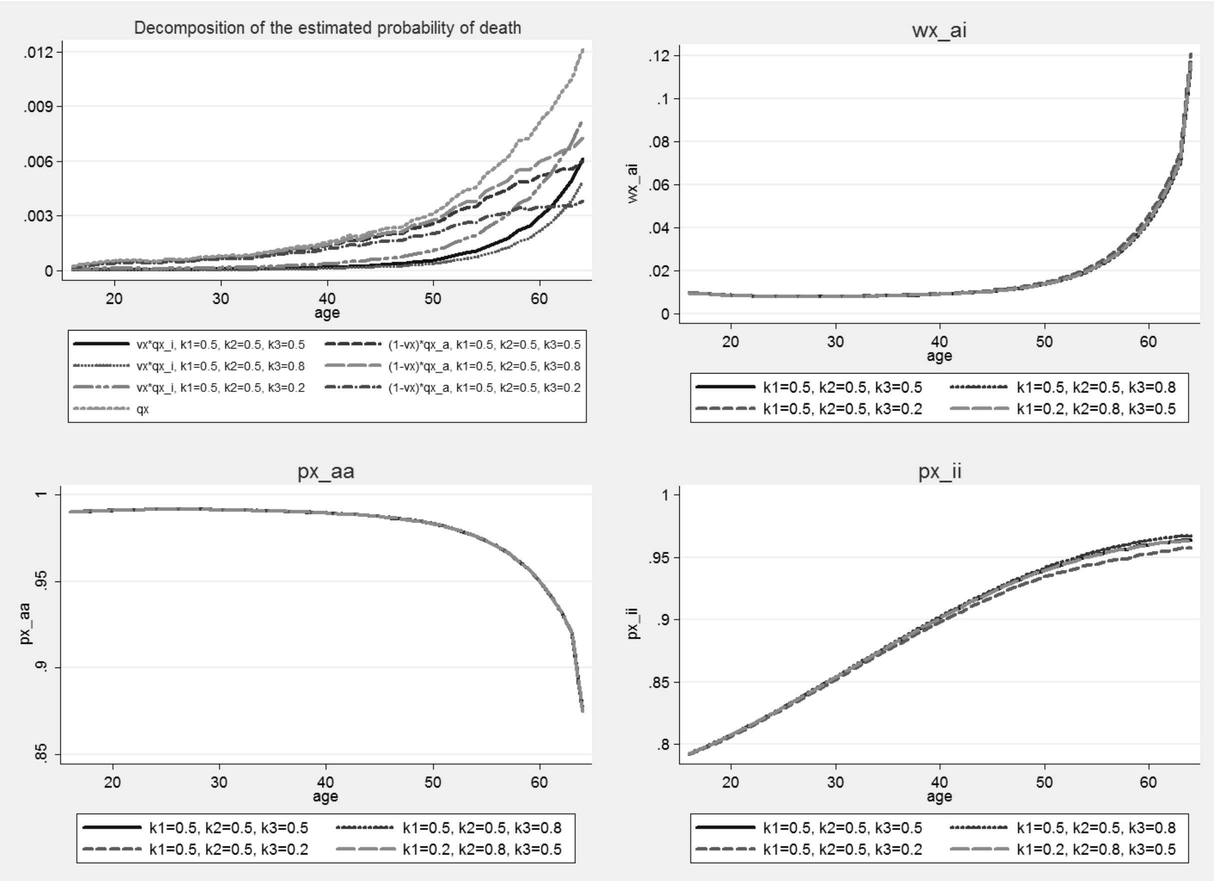
Additionally, the values of q_x^{i} in all scenarios are almost exactly equal to the values of q_x^{ii}, whereas the values of q_x^{ia} are extremely small, i.e. nearly zero. It means the probability that an inactive person dies in the same state (q_x^{ii}) is likely to be the main component of the probability of dying of inactive people (q_x^{i}). As expected, in the case of an active individual, the probability of dying for active people aged x (q_x^{a}) is also mostly determined by the probability of dying when individuals are still in an active state (q_x^{aa}).

Figure 7: Estimates of annual probabilities of death with any various scenarios of k_1, k_2 and k_3 for male in 2011



Source: The authors' own calculation based on the one-year multiple state model

Figure 8: Decomposition of the estimated probability of death, the probability of becoming inactive and the probability of surviving in the same state with various scenarios of k_1, k_2 and k_3 for male in 2011



Source: The authors' own calculation based on the one-year multiple state model

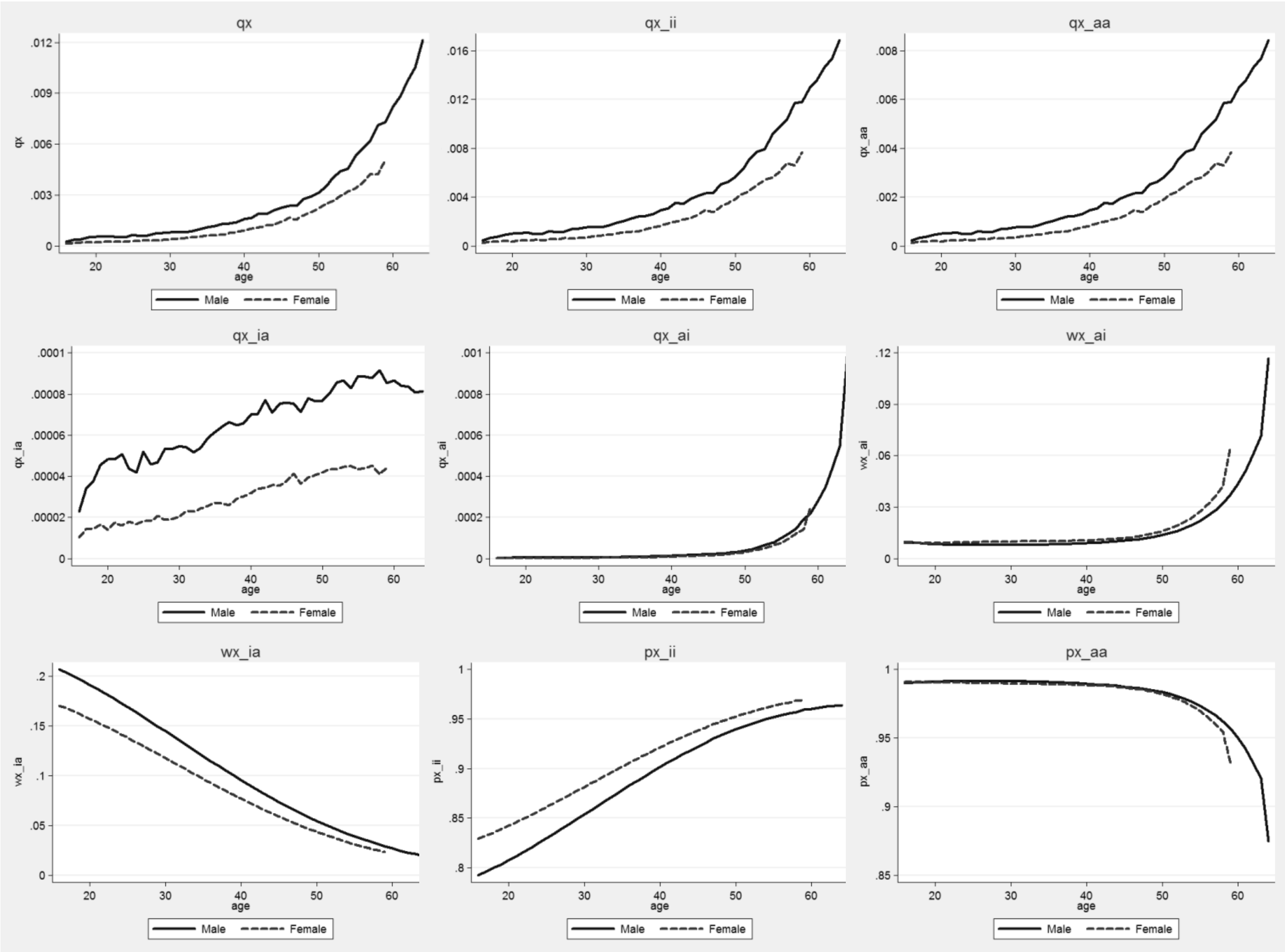
Figure 8 plots the decomposition of the estimated probability of death between the contribution of $v_x q_x^i$ and $(1-v_x)q_x^a$, the probabilities of becoming inactive (w_x^{ai}) and the probabilities of surviving in the same state over one year p_x^{aa} and p_x^{ii} with the different scenarios of k_1, k_2 and k_3. The graphic of the decomposition of the estimated probability of death shows that the annual mortality rate (q_x) is largely affected by the mortality of the active people $(1-v_x)q_x^a$ with the higher value of k_3. For example, the value of $k_3 = 0.8$ produces the highest value of $(1-v_x)q_x^a$ and the lowest value of $v_x q_x^i$.

The probability of becoming inactive within one year (w_x^{ai}) increases with age and steeply rises over the age of 50, in contrast to the probability of p_x^{aa}. Furthermore, the probability of p_x^{ii} rises continuously with age and is associated adversely with the probability of recovering from an inactive to an active state within one year (w_x^{ia}).

Similarly to the gender differences in general mortality rates q_x, the probabilities of death in any state of a person aged x, i.e. $q_x^{ii}, q_x^{aa}, q_x^{ia}$ and q_x^{ai} for men are greater than women, as shown in Figure 9. The probabilities of death in the same state q_x^{ii} and q_x^{aa} for both genders have the same pattern as the mortality rate q_x increasing with age, but the inactive people have a higher probability of death than for both general population and active people. Moreover, it is still rare for anyone to die in a different state within one year of a transition; as a result, the probabilities of inactive (or active) people dying in the different state q_x^{ia} (or q_x^{ai}) are almost zero.

The probabilities of becoming inactive (w_x^{ai}) are increasing with age and are higher among females than males. On the other hand, the probabilities of recovering from an inactive to an active state (w_x^{ia}) have been decreasing with age and are lower for females than males. We also compare the probability of surviving in the same state p_x^{ii} and p_x^{aa} for both genders. The probabilities that the inactive people are still inactive (p_x^{ii}) rise with age and there is a higher rate of inactive females who are still inactive than males. In contrast, the probabilities of active people being in the active state (p_x^{aa}) decrease with age and the active males have more chance to stay in the same state than females.

Figure 9: Estimates of annual probabilities in any state with $k_1 = 0.5, k_2 = 0.5, k_3 = 0.5$ by age and gender in 2011



Source: The authors' own calculation based on the one-year multiple state model

5. Conclusion

This paper proposes the one-year discrete time multiple state model of working-age disabled people using the self-reported cross-sectional disability data. We also allow for the recovery from an inactive to an active state in the model, whereas previous research focuses on the elderly and does not consider their recovery. The disability prevalence rates, the mortality rates for the general population and the assumptions regarding the relative mortality ratio between non-disabled and disabled individuals are used to estimate the state probabilities and transition probabilities between states.

The estimated gender- and age-specific disability prevalence rates that represent the probability of being inactive increase with age and are greater among women than men, whereas the one-year recovery rates as a proxy of the probability of recovery from an inactive to an active state decrease with age and men recover their health and get back to work at a greater rate than women. The probabilities of becoming inactive are nearly equal for younger ages and then rise rapidly at older ages. Moreover, the size of the relative mortality ratio among active and inactive people remaining in the same state is the main determinant of the probabilities of dying in the same state in one year, q_x^{ii} and q_x^{aa}. On the other hand, the probabilities of death in the different state, q_x^{ai} and q_x^{ia}, hardly occur, i.e. are almost zero. Consequently, the probabilities of death among each group, q_x^{i} and q_x^{a} would be approximated by the probability of dying in the same state.

The model framework presented in this paper is applicable when the disability prevalence rates are available. However, the disability rates might be replaced with the other prevalence measures as indicators of long-term health problems e.g. activities living daily (ADL) or instrumental activities of daily living (IADL). Our proposed method could be applied to project the size of the different groups, i.e. active, inactive and dead, and to evaluate the demand for the incapacity benefits. In a future study, we will focus on using these estimated transition probabilities to measure the future cost of government spending on disability benefits.

Appendix A

One-year transition probabilities

- An active individual

The fundamental relations of one-year probabilities and transition probabilities of an active individual age x are defined as follows:

$$p_x^{aa} + p_x^{ai} = p_x^{a} \tag{A.1}$$

$$q_x^{aa} + q_x^{ai} = q_x^{a} \tag{A.2}$$

$$p_x^{a} + q_x^{a} = 1 \tag{A.3}$$

$$p_x^{ai} + q_x^{ai} = w_x^{ai} \tag{A.4}$$

$$p_x^{aa} + q_x^{aa} = 1 - w_x^{ai} \tag{A.5}$$

where

p_x^{aa} is the probability that an active person aged x is alive in an active state at age $x+1$

q_x^{aa} is the probability that an active person aged x dies in an active state at age $x+1$.

p_x^{ai} is the probability that an active person aged x is alive in an inactive state at age $x+1$

q_x^{ai} is the probability that an active person aged x dies in an inactive state at age $x+1$.

p_x^{a} is the probability that an active person aged x is alive at age $x+1$.

q_x^{a} is the probability that an active person aged x dies within one year.

w_x^{ai} is the probability that an active person aged x becomes inactive within one year.

- An inactive individual

The one-year conditional probabilities related to an inactive individual age x are hold in the following relations:

$$p_x^{ii} + p_x^{ia} = p_x^{i} \quad \text{(A.6)}$$

$$q_x^{ii} + q_x^{ia} = q_x^{i} \quad \text{(A.7)}$$

$$p_x^{i} + q_x^{i} = 1 \quad \text{(A.8)}$$

$$p_x^{ia} + q_x^{ia} = w_x^{ia} \quad \text{(A.9)}$$

$$p_x^{ii} + q_x^{ii} = 1 - w_x^{ia} \quad \text{(A.10)}$$

where

p_x^{ii} is the probability that an inactive person aged x is alive in an inactive state at age $x+1$

q_x^{ii} is the probability that an inactive person aged x dies in an inactive state at age $x+1$

p_x^{ia} is the probability that an inactive person aged x is alive in an active state at age $x+1$

q_x^{ia} is the probability that an inactive person aged x dies in an active state at age $x+1$

p_x^{i} is the probability that an inactive person aged x is alive at age $x+1$

q_x^{i} is the probability that an inactive person aged x dies within one year

w_x^{ia} is the probability that an inactive person aged x recover to an active state within one year.

Appendix B

The parabolic function of q_x^{ii} is expressed as:

$$A\left(q_x^{ii}\right)^2 + Bq_x^{ii} - q_x = 0 \quad \text{(A.11)}$$

where $A = -k_1 k_3 \left[\left(1 - v_x\right) + v_x k_1 k_2 w_x^{ia} \right]$

and $B = k_3 \left(1 - v_x + v_x k_2 w_x^{ia}\right) + k_1 v_{x+1} \left(1 - q_x\right) + k_1 \left(q_x - v_x + v_x w_x^{ia}\right) + v_x$

Following the quadratic formula to solve the equation (A.11), there are two real solutions with the positive values. However, we obtain the unique solution that exists the (0,1) interval as follows:

$$q_x^{ii} = \frac{-B + \sqrt{B^2 + 4Aq_x}}{2A} \qquad \text{(A.12)}$$

References

Albarran, I., Ayuso, M., Guillén, M. and Monteverde, M. (2005) A multiple state model for disability using the decomposition of death probabilities and cross-sectional data. *Communications in Statistics: Theory and Methods*, **34**(9/10), 2063-2075.

Bahk, J., Kang, H. Y., & Khang, Y. H. (2019) The Life Expectancy Gap between Registered Disabled and Non-Disabled People in Korea from 2004 to 2017. *International journal of environmental research and public health*, **16**(14), 2593.

Bell, B. and Smith, J. (2004) *Health, disability insurance and labour force participation*. Working Paper No. 218, Bank of England, London.

De Jong, P. and Heller, G. Z. (2008) *Generalized Linear Models for Insurance Data*. Cambridge: Cambridge University Press.

Fong, J. H., Shao, A. W. and Sherris, M. (2015) Multistate actuarial models of functional disability. *North American Actuarial Journal*, **19**(1), 41-59.

Forman-Hoffman, V. L., Ault, K. L., Anderson, W. L., Weiner, J. M., Stevens, A., Campbell, V. A., & Armour, B. S. (2015) Disability status, mortality, and leading causes of death in the United States community population. *Medical care*, **53**(4), 346–354.

Frees, E. W. (2009) *Regression Modeling with Actuarial and Financial Applications*, Cambridge University Press.

Guillen, M. (2014) Regression with categorical dependent variables. In *Predictive Modeling Applications in Actuarial Science Volume I: Predictive Modeling Techniques* (eds. Frees, E. W., Derrig, R. A. and Meyers, G.), Cambridge University Press.

Haberman, S. and Pitacco, E. (1999) *Actuarial Models for Disability Insurance*. Chapman & Hall/CRC.

Hariyanto, E. A., Dickson, D. C. M. and Pitt, D. G. W. (2014) Estimation of disability transition probabilities in Australia I: Preliminary. *Annals of Actuarial Science*, **8**(1), 131-155.

Hosmer, D. W., Lemeshow, S. and Sturdivant, R. X. (2013) *Applied Logistic Regression*. Third ed., New York: Wiley.

Howse, K. (2006) *Life expectancy and the compression of morbidity: A critical review of the debate*. Oxford Institute of Ageing Working Paper 226.

Jones, M. K. (2008) Disability and the labour market: A review of the empirical evidence. *Journal of Economic Studies*, **35**(5-6), 405-424.

Lee, R. D. and Carter, L. R. (1992) Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association*, **87**(419), 659-671.

Leung, E. (2004) Projecting the needs and costs of long term care in Australia. *Australian Actuarial Journal*, **10(**2), 343-385.

Leung, E. (2006) A multiple state model for pricing and reserving private long term care insurance contracts in Australia. *Australian Actuarial Journal*, **12**(2), 187-247.

Majer, I. M., Nusselder, W. J., Mackenbach, J. P., Klijs, B., & van Baal, P. H. (2011) Mortality risk associated with disability: a population-based record linkage study. *American journal of public health*, **101**(12), e9–e15.

Majer, I., Stevens, R., Nusselder, W., Mackenbach, J. and van Baal, P. M. (2013) Modeling and forecasting health expectancy: Theoretical framework and application. *Demography*, **50**(2), 673-697.

McVicar, D. (2008) Why have UK disability benefit rolls grown so much?. *Journal of Economic Surveys*, **22**(1), 114-139.

Nations, U. (1968) *The concept of a stable population - application to the study of populations of countries with incomplete demographic statistics*. United Nations.

Nuttall, S. R., Blackwood, R. J. L., Bussell, B. M. H., Cliff, J. P., Cornall, M. J., Cowley, A., Gatenby, P. L. and Webber, J. M. (1994) Financing long-term care in Great Britain. *Journal of the Institute of Actuaries (1886-1994)*, **121**(1), 1.

O'Reilly, A. (2007) *The right to decent work of persons with disabilities*, Geneva, Switzerland : International Labour Office.

OECD (2009) Pathways onto (and off) disability benefits: Assessing the role of policy and individual circumstances. In *OECD Employment Outlook 2009 : Tackling the Jobs Crisis*. OECD Publishing.

OECD (2010) *Sickness, Disability and Work: Breaking the Barriers: A Synthesis of Findings across OECD Countries*. OECD Publishing.

OECD (2014) *Social expenditure update-social spending is falling in some countries, but in many others it remains at historically high levels.*

OECD (2017) *Public spending on incapacity (indicator)*. doi: 10.1787/f35b71ed-en (Accessed on 18 January 2017).

Pitacco, E. (2014) *Health Insurance : Basic Actuarial Models*. Springer International Publishing.

Renshaw, A. E. and Haberman, S. (1995) On the graduations associated with a multiple state model for permanent health insurance. *Insurance: Mathematics and Economics*, **17**(1), 1-17.

Renshaw, A. E. and Haberman, S. (2000) Modelling the recent time trends in UK permanent health insurance recovery, mortality and claim inception transition intensities. *Insurance Mathematics and Economics*, **27**, 365-396.

Rickayzen, B. D. and Walsh, D. E. P. (2002) A multi-state model of disability for the United Kingdom: Implications for future need for long-term care for the elderly. *British Actuarial Journal*, **8**(2), 341-393.

Soubbotina, T. P. and Sheram, K. (2000) *Beyond economic growth : meeting the challenges of global development*. World Bank, Washington, D.C.

United Nations (1968) *The concept of a stable population - application to the study of populations of countries with incomplete demographic statistics*, United Nations Publications.

Webber, D. A. and Bjelland, M. J. (2015) The impact of work-limiting disability on labor force participation. *Health Economics*, **24**(3), 333-352.

WHO (2011) *Global Health and Aging*. Geneva: World Health Organization (WHO).