Disability and survival among people aged 50+: the English Longitudinal Study of Ageing

Marjan Qazvini *

Abstract

This study aims to investigate the association between disability and noncommunicable diseases (NCDs) with survival among people aged 50 and older. We use data from five waves of the English Longitudinal Study of Ageing (ELSA) and apply discrete time-to-event (survival) analysis using the generalised linear model (GLM) and the generalised linear mixed model (GLMM). Our results confirm gender differences in health and survival and show that mobility and instrumental activities of daily living (IADL) disabilities and NCDs such as cancer significantly affect survival in old-aged population.

1 Introduction

The ageing population is one of the challenges that some countries are currently dealing with and many more will be dealing with in the future. After retirement, very few individuals are still working and contribute to the economy and society. The majority of old people start developing noncommunicable diseases (NCD) and difficulties in performing Activities of Daily Livings (ADL) and Instrumental ADL (IADL). According to WHO, about 15% of the world's population have some form of disability and among the adults between 110 million and 190 million have significant difficulties in functioning. Population ageing and the rise of chronic diseases are two factors that contribute to the increase in the rate of disability. [1]. In 2015 WHO reported that 40 million out of 56 million global deaths were due to NCDs. [2]

In this study, we aim at investigating the impact of disability and NCDs on survival among population aged 50+. The International Classification of Functioning, Disability and Health (ICF) defines disability as impairments, activity limitations and participation restrictions. Here, we consider disability as mobility, ADL and IADL impairment.

*Department of Actuarial Mathematics and Statistics, School of Mathematical and Computer Sciences, Heriot-Watt University Malaysia, M.Qazvini@hw.ac.uk

[1]https://www.who.int/news-room/fact-sheets/detail/disability-and-health

[2]https://www.who.int/gho/ncd/en/

We will show how the effect is different among men and women. For this, we apply discrete time-to-event, also known as survival analysis, to the English Longitudinal Study of Ageing (ELSA), which is a panel study of a representative cohort of individuals aged 50 or older living in private households in England. ELSA has been extensively studied by researchers in medical and social sciences. For example, Demakakos et al. (2015) study the relationship between wealth and all-cause and cause-specific mortality and find that wealth is strongly associated with cardiovascular disease (CVD) and other non-cancer mortality among people aged 50-64, whereas there is a weak relationship between wealth and cancer mortality. Kessler et al. (2020) focus on the comparison of risk factors such as physical inactivity, smoking, hypertension, diabetes and high BMI which are the causes of NCD among the population aged 60+ from ELSA and Bagé Cohort study of Ageing (SIGa-Bagé). They conclude that the level of these risks among the Brazilian population is higher than the English population. They explain their results by referring to healthcare and economic situations in England. Some of the studies in social sciences consider the impact of social interaction and well-being on survival among people in old age. Steptoe et al. (2015) investigate the link between psychological wellbeing and health among old-aged people. They find that illnesses such as CVD and chronic lung disease (CLD) give rise to a reduced level of life satisfaction and having a purpose in life is associated with longer survival. Khondoker et al. (2017) investigate the relevance of social relationships for cognitive health among people in old age. Their findings show that positive support from children is associated with reduced risk of dementia, whereas negative social support regardless of the type of relationship is associated with increased risk of dementia. In another study Rofnsson et al. (2017) consider the relationship between social relationships and dementia and find that there is a positive association between loneliness and developing dementia.

The changes in disability prevalence over time have been considered in the literature. Lin et al. (2012) study the US population aged 70+ between the years 1982 and 2009. They adopt two models: an unadjusted and a model adjusted for age, period, cohort and sociodemographic factors. They find that IADL disability has decreased, but ADL disability has been stable. They conclude that US old-aged population of recent cohorts have become more disabled. In a study of the German population aged 50-79, Buttery et al. (2016) consider the trends of physical functioning over time. They observe an improvement in physical functioning for both men and women with the results being less pronounced for men aged 65-79. They justify their results by examining different health behaviours among men and women over time. Pongiglione et al. (2016) investigate the association between disability and survival among English old-aged population. They consider three types of disability viz. impairment in eyesight/hearing and chronic conditions, activity limitations (ADLs) and participation restrictions (IADLs) and carry out a discrete-time survival analysis. They show that even though increased severity of the disability is associated with a reduced level of survival for both men and women, women have more disability problems than men. They also find that such association decreases over time for men, but remains constant for women. This suggests men can become more resilient to disability over time.

In another study, Pongiglione et al. (2017a) examine the changes in disability rate among the English population over the period 2002-2012 according to 4 classes viz. no disability, mild, moderate and severe. They find that over this period, fewer women suffer from severe and moderate disability and mild disability has increased. On the other hand, severe disability has not changed for men and that more men suffer from a moderate disability. Binary and multi-classification of disability have been examined by Pongiglione et al. (2017b). According to their findings the most informative model is a 4-class model which includes *no*, *mild*, *moderate* and *severe* disability.

In this study, we consider the self-report binary classification of disability and we measure the severity by the number of activities limitation reported by participants. We perform discrete time-to-event analysis to see how survival among men and women is affected by disability and chronic diseases. This paper is organised as follows: In Section 2 we explain our data and data preparation in a panel study and then in Section 3 we provide a review of discrete-time survival models. In Section 4 we present our results followed by conclusion and discussion in Section 5.

2 Data and data preparation

The English Longitudinal Study of Ageing (ELSA) is a collection of economics, social, psychological, cognitive, health, biological and genetic data. The study commenced in 2002 and the sample has been followed up every 2 years. The first cohort was selected from respondents to the Health Survey for England (HSE) in 1998, 1999 and 2001 and included people borne on or before February 29, 1952, i.e. aged 50 and older. The first ELSA wave was in 2002-2003. Wave 2 took place in 2004-2005, wave 3 in 2006-2007, wave 4 in 2008-2009 and wave 5 in 2010-2011. To make sure ELSA is designed to be representative of people aged 50 and over in England, at waves 3 and 4, a refreshment cohort of people just entering their 50s has been introduced. At wave 2, an End of Life interview was conducted with the purpose of finding out about the health and socio-economic situations of people just before their death. End of Life interviews have been carried out at waves 2, 3, 4 and 6. There was no End of Life interview at wave 5. Until now 9 waves have been completed. However, in ELSA documentation the results of only 8 waves are available. (For more information on ELSA, sampling and interview process see, for example, Steptoe et al. (2013) and Blake et al. (2015)). Table 1 shows the number of participants in each cohort and the number of deaths among core members reported in waves 2, 3, 4 and 6.

Variables

In this study, we extract information about age, gender, employment status, marital status, self-rated physical and health conditions of the first cohort of core members. As it is normal with survey data we have the problem of attrition. This problem arises, for example, when individuals drop out after the first wave and rejoin in later

Table 1: The number of core members and deaths in each wave

Waves	Cohort 1	Cohort 3	Cohort 4	Death$^{(2)}$
1) 2002-2003	11,391	-	-	-
2) 2004-2005	$8,780^{(1)}$	-	-	133
3) 2006-2007	7,535	1,275	-	369
4) 2008-2009	6,623	972	2,291	234
5) 2010-2011	6,242	936	1,912	-
6) 2012-2013	-	-	-	240

NatCen: Technical Report (Wave 6)
1- In wave 2 one member aged below 50 has been removed
2- The number of deaths of the core members

waves. We collect information of core members from waves 1, 2, 3, 4 and 5 and retrieve information regarding their status from waves 2, 3, 4 and 6. The variables that we consider in our study are presented in Table 2. These variables are extracted from "core data" files that can be obtained from ELSA website after registration. The participants have been asked to fill in a questionnaire. Information provided is based on participants' self-assessment of their health conditions. There are some categories of responses such as *refusal*, *do not know* or *not applicable* that we consider as missing values. Our dependent variable is the *status* of the individuals that can be obtained from "eol" files for different waves. The *age* is a continuous variable which is limited to 90 to avoid disclosure. There are 5 categories of *marital status*. We combine the categories of civil partnership with marital status. For example, *separation after civil partnership* or *death of one of the partners* are combined under *divorced* and *widowed*, respectively. There are 5 categories of *employment status*. We set all missing values for age 60+ as *retired*. The category *other* includes *self-employed*, *looking after home or family*, *unemployed* and missing values. The variables *mobility*, *ADL* and *IADL* score are discrete. They represent the number of activities with which participants have difficulties or need help. For example, score "0" means that the participants have not reported difficulties in performing any activities related to mobility and score "10" means they have reported difficulties in performing all 10 activities. ADL and IADL scores are defined, similarly. The *NCDs* in this study are provided in Table 2. At each wave, we consider a newly-diagnosed disease. That is the respondents have been asked whether they have recently experienced the symptoms or they have been told by their doctors that they had a specific disease. Table 3 shows the proportion of deaths (1) and survivors (0) in different age groups, i.e. middle-age $(50-65)$, young-old $(65-75)$, old $(75-85)$ and old-old (85+) among men and women. For all ages, we can observe an increasing trend until age 85 and after that, the proportion of people who die decreases. We can also observe that the proportion of women who die before age 85 is less than men, which agrees with our expectation that women have higher life expectancy than men. Figure 1 shows the marital status in different waves. We can see a similar pattern in all waves. A large proportion of participants are married or are in any other form of partnership. The next largest category is widowed. The categories divorced and remarried are almost equal in all waves and singles form the smallest group. Figure

Table 2: The dependent and independent variables of the study

Age	Continuous	age above 90 is recorded as 90 to avoid disclosure $(50 - 90)$
Gender	Categorical	Male, female
Marital status	Categorical	Single, married, divorced, remarried, widowed
Employment status	Categorical	Employed, retired, permanently sick or disabled, unemployed, other (self-employed, looking after home or family, unemployed and NAs).
Mobility	Discrete $0, \dots, 10$	Walking 100 yards, sitting for about two hours, getting up from chair, climbing several flights of stairs without resting, stooping, kneeling, reaching or extending your arms, pulling or pushing large objects, lifting or carrying weights picking up a 5p coin
ADL	Discrete $0, \dots, 6$	Dressing, including putting on shoes and socks walking across a room bathing or showering eating, such as cutting up your food getting in or out of bed using toilet
IADL	Discrete $0, \dots, 7$	Using a map preparing a hot meal shopping for groceries making telephone calls taking medications doing work around the house or gardens managing money, such as paying bills
NCD	Yes No	High blood pressure, heart attack, diabetes, stroke Arthritis, chronic lung disease, cancer, Parkinson's, Alzheimer and dementia.
Status	Alive Dead	Dependent variable

2 shows employment status in 5 waves. Most participants are retired, which is not surprising given the age range in our study. The next largest category is employed and the smallest one is unemployed.

Figure 3 presents a list of activities that participants have been asked whether they had difficulty in performing them or not. The blue dots represent activities related to ADL, the red dots activities related to IADL and the green ones activities related to Mobility. We can learn from this figure (i) very few people have reported difficulty with "walking 100 yards"; (ii) mobility disability is more common than ADL and IADL disabilities - more than half of reported difficulties are related to mobility; (iii) IADL disability of "taking medications" and "making telephone calls" are less common than other disabilities.

Tables 4, 5 and 6 provide the percentage of participants with different levels of mobility, ADL and IADL disability in different age groups, respectively. In Tables 4 and 5 the percentage of participants with disability decreases as the score increases. Also, we can see that in Table 4 the percentage of participants with disability decreases

Table 3: The number of deaths and survivors in different age groups (%)

	Alive/Dead	Middle-age 50-65	Young-old 65-75	Old 75,85	Old-old 85+	Total
Female	0	27.8	14.7	8.3	2.3	53.1
	1	0.3	0.5	0.8	0.4	2.0
Total		28.1	15.2	9.1	2.7	55.1
Male	0	23.1	12.4	5.8	1.1	42.4
	1	0.5	0.8	1.0	0.3	2.6
Total		23.6	13.2	6.8	1.4	45.0
All	0	50.9	27.1	14.1	3.4	95.5
	1	0.8	1.3	1.8	0.7	4.6
Total		51.7	28.4	15.9	4.1	100

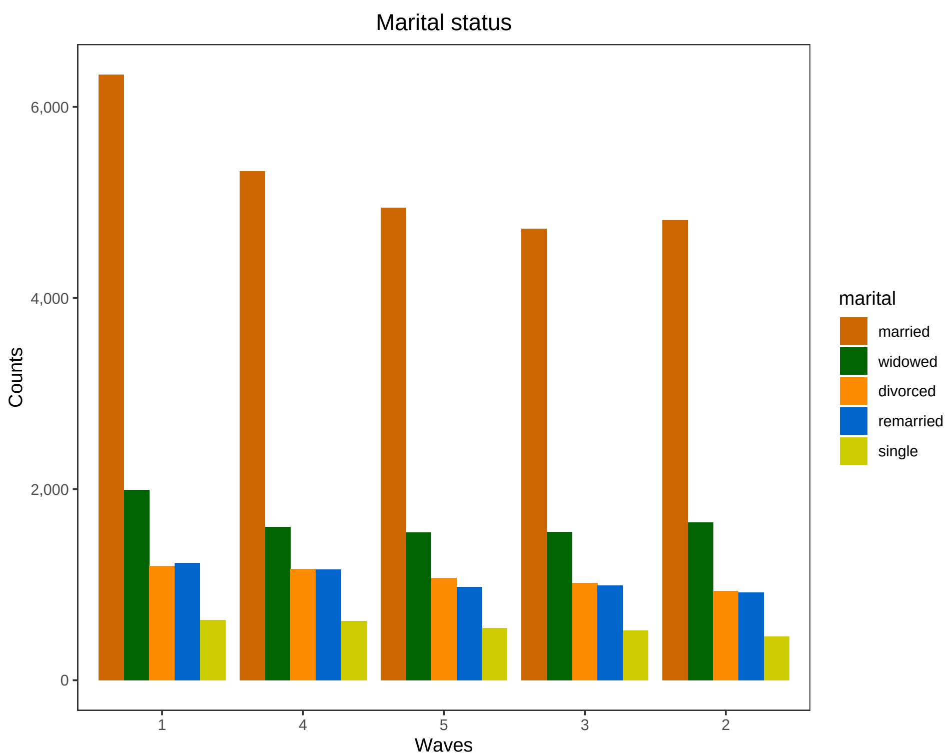


Figure 1: Marital status in different waves

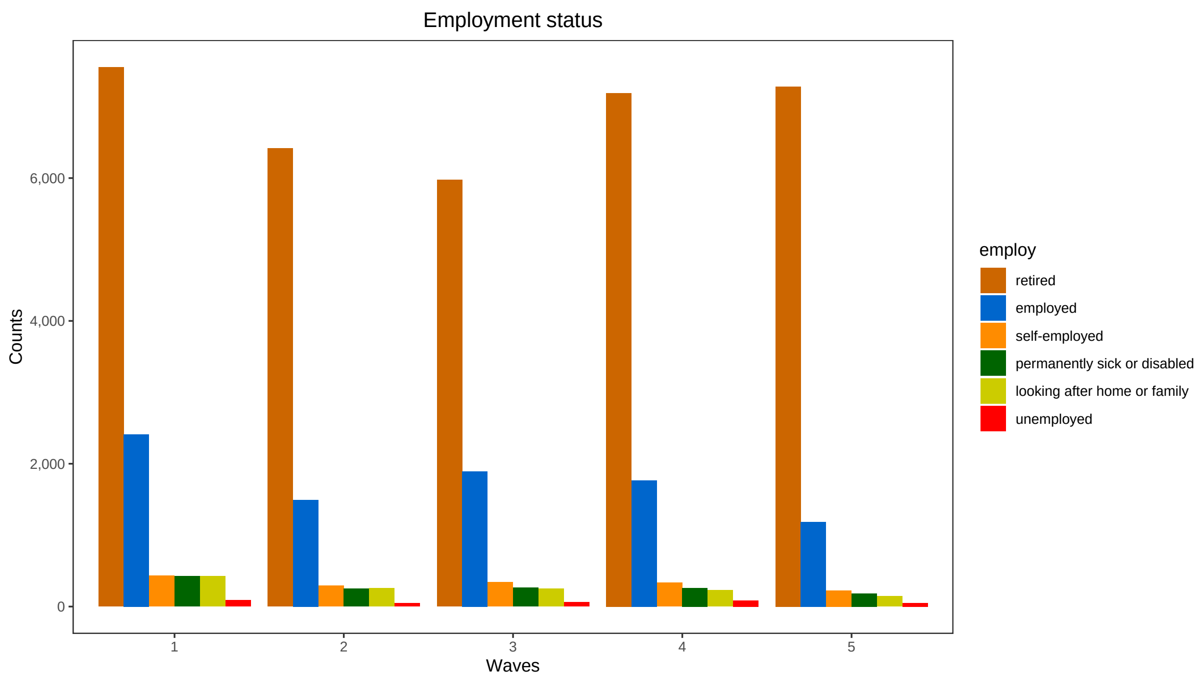


Figure 2: Employment status in different waves

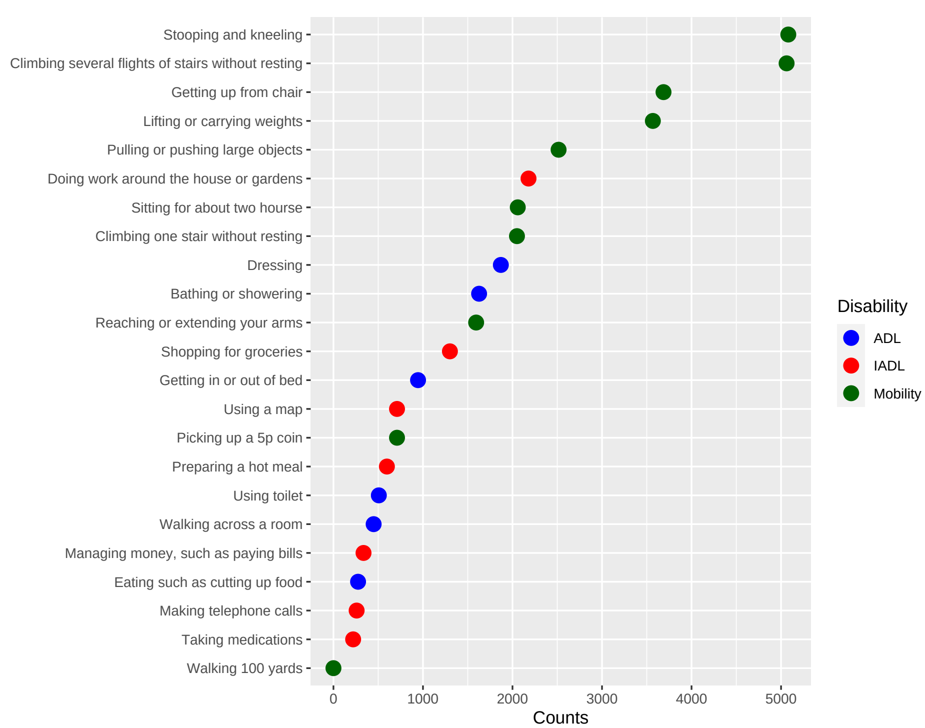


Figure 3: The number of participants with difficulties in performing different activities

Table 4: Mobility score according to different age group (%)

	0	1	2	3	4	5	6	7	8	9	10
Middle-age	27.5	8.3	4.5	3.0	2.1	1.6	1.3	1.1	1.0	0.9	0.4
Young-old	10.8	5.0	3.5	2.4	1.8	1.3	1.2	1.0	0.7	0.5	0.2
Old	3.8	2.5	2.0	1.7	1.4	1.2	1.0	0.9	0.8	0.4	0.2
Old-old	0.5	0.4	0.4	0.4	0.4	0.4	0.3	0.5	0.4	0.2	0.1
Total	42.6	16.2	10.4	7.5	5.7	4.5	3.8	3.5	2.9	2.0	0.9

Table 5: ADL score according to different age group (%)

	0	1	2	3	4	5	6
Middle-age	44.7	3.4	1.5	0.9	0.6	0.4	0.2
Young-old	22.4	3.3	1.3	0.7	0.4	0.2	0.1
Old	10.7	2.7	1.2	0.6	0.3	0.2	0.1
Old-old	2.1	0.8	0.5	0.2	0.2	0.1	0.2
Total	79.9	10.2	4.5	2.4	1.5	0.9	0.6

as they get older. It is worth noting that some of the activities related to mobility, ADL and IADL overlap. For example, if an individual has difficulty in “lifting or carrying weights”, it is very likely that they have also difficulty in “shopping for groceries”. In Table 5 we can see that the percentage of participants with score of 6 is either constant or increases as they age. Similar pattern can be observed in Table 6, where the percentage of participants with score of 7 is higher in the old-old group. From all these tables we can conclude that IADL disabilities are more common in the old-old group. This is also illustrated in Figure 4, where we can observe an increasing trend in age for all types of disabilities with the slope being steeper for IADL disability.

3 Models

In this section we look at the discrete-time survival models, also known as time-to-event models, as an approximation to continuous-time survival models. When we are working with survey data the information is not available at the exact point in time and we only know the period in which the event of interest occurs, which is usually every one year or in our case every two years. Discrete time-to-event models consider the change in categorical dependent variable Y in a time interval. We can divide the underlying continuous-time process into intervals $[0, a_1)$, $[a_1, a_2)$, $\ldots$, $[a_{t-1}, a_t)$, $[a_t, \infty)$. Let T be a discrete-time random variable, where $T = t$ means the event has happened in the

Table 6: IADL score according to different age group (%)

	0	1	2	3	4	5	6	7
Middle-age	44.2	4.0	1.9	1.0	0.3	0.1	0.1	0.1
Young-old	22.6	3.1	1.4	0.7	0.3	0.2	0.1	0.1
Old	10.1	2.8	1.5	0.6	0.3	0.2	0.2	0.2
Old-old	1.6	0.7	0.6	0.4	0.3	0.2	0.1	0.3
Total	78.5	10.6	5.4	2.7	1.2	0.7	0.5	0.7

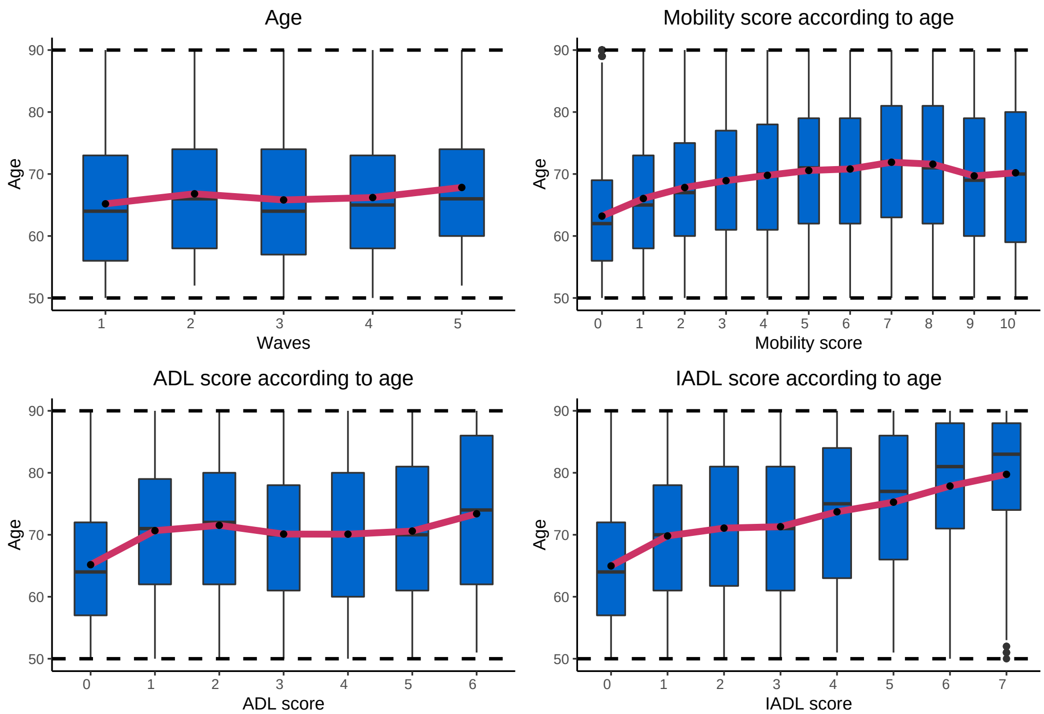


Figure 4: The distribution of age according to different scores

interval $[a_{t-1}, a_t)$ and $\mathbf{x}$ be a vector of covariates. We can now define the discrete-time hazard function as the probability that an individual i experiences the event of interest in period T given that this event has not happened before time t. For individual i, the discrete-time hazard function is given by

$$\lambda_i(t|\mathbf{x}) = \Pr(T = t | T \geq t, \mathbf{x}).$$

Another related definition is the survival probability, i.e. the probability that an individual does not experience an event before time t, which in discrete-time is given by

$$S_i(t|\mathbf{x}) = \Pr(T > t|\mathbf{x}) = \prod_{s=1}^{t} (1 - \lambda_i(s|\mathbf{x})).$$

By introducing explanatory variables to our model, we can write the hazard function as

$$\lambda_i(t|\mathbf{x}) = G(\beta_0(t) + \beta_1 x_{1it} + \cdots + \beta_k x_{kit} + \gamma_1 z_{1i} + \cdots + \gamma_j z_{ji})$$

where $(x_{1it}, \ldots, x_{kit})$ are k time-varying independent variables such as change of employment status and $(z_{1i}, \ldots, z_{ji})$ are j time-constant variables such as sex and permanently sick status. In this model, $\beta_1, \ldots, \beta_k$ and $\gamma_1, \ldots, \gamma_j$ are corresponding regression coefficients and G is a suitable link function to provide values of $\lambda_i(t)$ within the proper interval $[0, 1]$. The intercept may be a function of t to show trends in time and can take the linear, quadratic, exponential and discontinuous functions. The intercept can be interpreted as a baseline hazard, that is, the hazard always present for any given set of covariates. If $\beta_0(t) = \beta_0$ there is no time trend and we have a simple regression model. In a survey study, when we have repeated measures for an individual, the independent assumptions of observations no longer holds. In this case, we use a random-intercept model, which is similar to fitting separate GLM for each individual with the same slope and different intercept. The suitable candidates for $G(.)$ are logistic and Gompertz also known as complementary log-log (clog-log) function. The corresponding hazard function for a logistic function is given by

$$\lambda_i(t|\mathbf{x}) = \frac{\exp(\beta_0(t) + \beta_1 x_{1it} + \cdots + \beta_k x_{kit} + \gamma_1 z_{1i} + \cdots + \gamma_j z_{ji})}{1 + \exp(\beta_0(t) + \beta_1 x_{1it} + \cdots + \beta_k x_{kit} + \gamma_1 z_{1i} + \cdots + \gamma_j z_{ji})}$$

and for a clog-log model by

$$\lambda_i(t|\mathbf{x}) = 1 - \exp\big(- \exp(\beta_0(t) + \beta_1 x_{1it} + \cdots + \beta_k x_{kit} + \gamma_1 z_{1i} + \cdots + \gamma_j z_{ji})\big).$$

Both these models can be readily applied in R. For more information on discrete time-to-event analysis see, for example, Austin (2017) or Tutz and Schmid (2016).

4 Results

In this section, we apply the discrete-time survival model to our data and compare our results for different functions such as logistic, cloglog and Gumbel. First of all, we need to convert our data set into long format. We can easily do this using Package *discSurv* [3] and *reshape* in R. In *discSurv*, there are three key inputs: exit times, status and unique id number for each individual. The output is a data frame with two new vectors that we use in function *glm*, *glmer* and/or *bglmer*. The new *status* vector is our dependent variable and the new *time interval* vector is used as one of our covariates. In *reshape*, first, we make sure that we have our data in wide format, i.e. repeated responses are in the same row and then convert this into long format. The *reshape* command can be used to change the format of our data. Here, the output is *time*, which can be used as one of our covariates, and *y*, which is our status. The next step is to remove information with lag 0, 4 and 3. For example, we eliminate those participants who were under observation only for one year, or who only participated in year 1 and year 5, i.e. the lag is 4, etc. The variance inflation factor is less than 5, therefore, there is no problem of multicollinearity. It is common for dichotomous variables to pose the separation problem. We solve this issue by setting all missing values for NCDs equal to zero. Since only newly-diagnosed diseases or disabilities are recorded under these variables, we assume that no entry is equivalent to no diseases or no disabilities.

[3]The command *dataLong* in this package prepares our data for regression analysis. In fact, we need to read our data in long format into R

Table 7: The estimates and p-values of the coefficients in three models

Covariates	Models					
	Clog-log		Logit		Gumbel	
	Estimate	p-value	Estimate	p-value	Estimate	p-value
Intercept	−14.8764	0.0000	−15.0865	0.0000	−3.9938	0.0000
Interval 2	3.5322	0.0000	3.5823	0.0000	0.7599	0.0000
Interval 3	3.4334	0.0000	3.4693	0.0000	0.7138	0.0000
Interval 4	3.1518	0.0000	3.1906	0.0000	0.6516	0.0000
Interval 5	2.3352	0.0000	2.3319	0.0000	0.4231	0.0000
Age	0.0867	0.0000	0.0893	0.0000	0.0241	0.0000
Sex (Male)	0.5548	0.0000	0.5895	0.0000	0.1579	0.0000
Marital status:						
Single	−0.6467	0.0066	−0.6687	0.0062	−0.1551	0.0095
Divorced	0.0285	0.8595	0.0398	0.8083	0.0377	0.3481
Remarried	−0.1171	0.4377	−0.1419	0.3687	−0.0456	0.2780
Widowed	−0.2706	0.0100	−0.2670	0.0151	−0.0532	0.0927
Employment status:						
Retired	0.7419	0.0281	0.6959	0.0403	0.0168	0.7741
Permanently sick or disabled	1.2689	0.0027	1.2283	0.0040	0.1358	0.1211
Unemployed	1.5429	0.0467	1.5418	0.0481	0.2510	0.1291
Other	0.9211	0.0450	0.9226	0.0452	0.1447	0.0756
Disability:						
Mobility score	0.0671	0.0005	0.0677	0.0008	0.0165	0.0046
ADL score	0.0660	0.0772	0.0784	0.0479	0.0257	0.0467
IADL score	0.1630	0.0000	0.1699	0.0000	0.0617	0.0000
NCDs:						
High blood pressure	0.0235	0.8454	0.0171	0.8923	0.0023	0.9494
Heart attack	0.3931	0.0642	0.4197	0.0652	0.1401	0.0564
Diabetes	0.2580	0.1455	0.2377	0.2086	0.0372	0.5181
Stroke	−0.1078	0.5650	−0.0317	0.8738	0.0679	0.3019
Arthritis	−0.1846	0.2315	−0.2032	0.2118	−0.0526	0.2481
Chronic lung disease	0.4265	0.0451	0.4490	0.0494	0.0961	0.1833
Cancer	1.1135	0.0000	1.2056	0.0000	0.3638	0.0000
Parkinson	0.3743	0.2584	0.5085	0.1743	0.1687	0.2519
Alzheimer	0.3678	0.1830	0.4755	0.1215	0.2369	0.0762
Dementia	0.2973	0.1178	0.4114	0.0496	0.2267	0.0079
AIC	5,340		5,331.8		5,328.9	

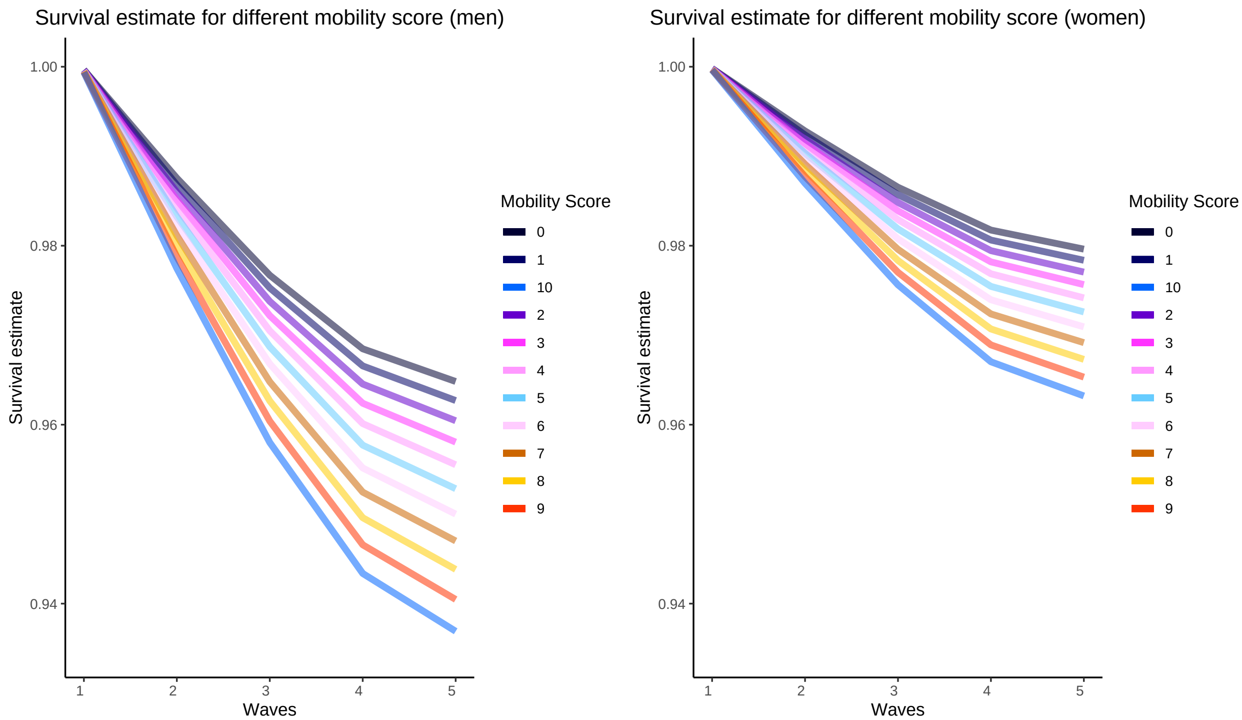


Figure 5: Survival estimate for different mobility score

The results for three link functions, namely, clog-log, logit and Gumbel are presented in Table 7. The first column shows our dependent variables used in all three models. The estimates obtained from logistic and clog-log models are very close. Positive coefficients imply that the corresponding covariates have a positive effect on hazards and hence a negative effect on survival. The intervals refer to the period $[a_{t-1}, a_t)$. Therefore, Interval 2 refers to the period between wave 1 and wave 2 $[1, 2)$, Interval 3 the period $[2, 3)$, Interval 4 the period $[4, 5)$ and finally Interval 5 the period $[5, 6)$. The coefficient of Intervals in all three models is positive and statistically significant. The coefficient of interval 5 is less than other intervals. We can explain this by referring to Figure 4. The plot on the top-left shows the distribution of age in different waves and we can observe that the average age in wave 5 is slightly higher than the average age in other waves. Age is a continuous random variable and has a positive coefficient. In the logistic model, we can say that 1-year increase in age leads to exp(0.089), i.e. 9.34% increase in the odds of risk of death. In the clog-log model, it is interpreted as exp(0.087), i.e. 9.06% increase in the risk of dying for every 1-year increase in age. This probability is about 2.4% for the Gumbel model. The coefficients of sex, mobility score, IADL score and cancer are positive and statistically significant in all models.
According to the coefficient of sex in the logistic model, the odds of dying among men is 80% higher than women. This can also be observed from Figures 5-8. In all these figures, the estimated survival curve for men decreases at a higher rate than women. In Figures 5 and 6 this pattern can be observed regardless of the health conditions and

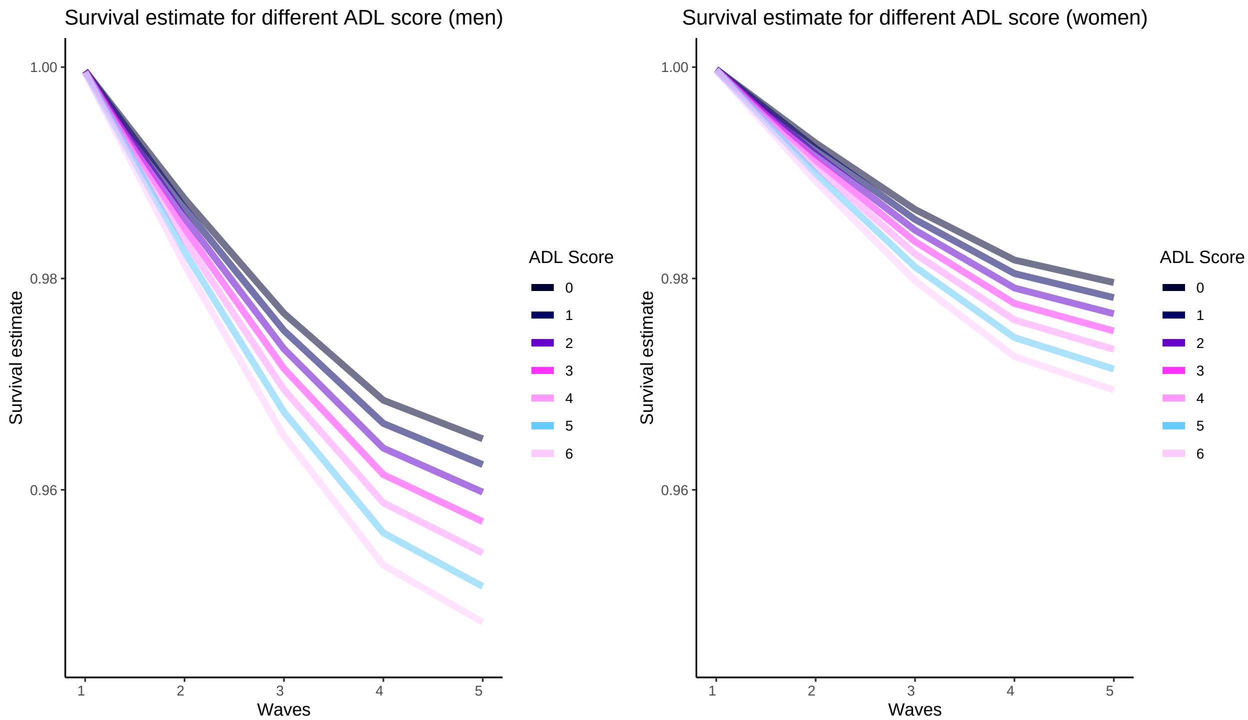


Figure 6: Survival estimate for different ADL score

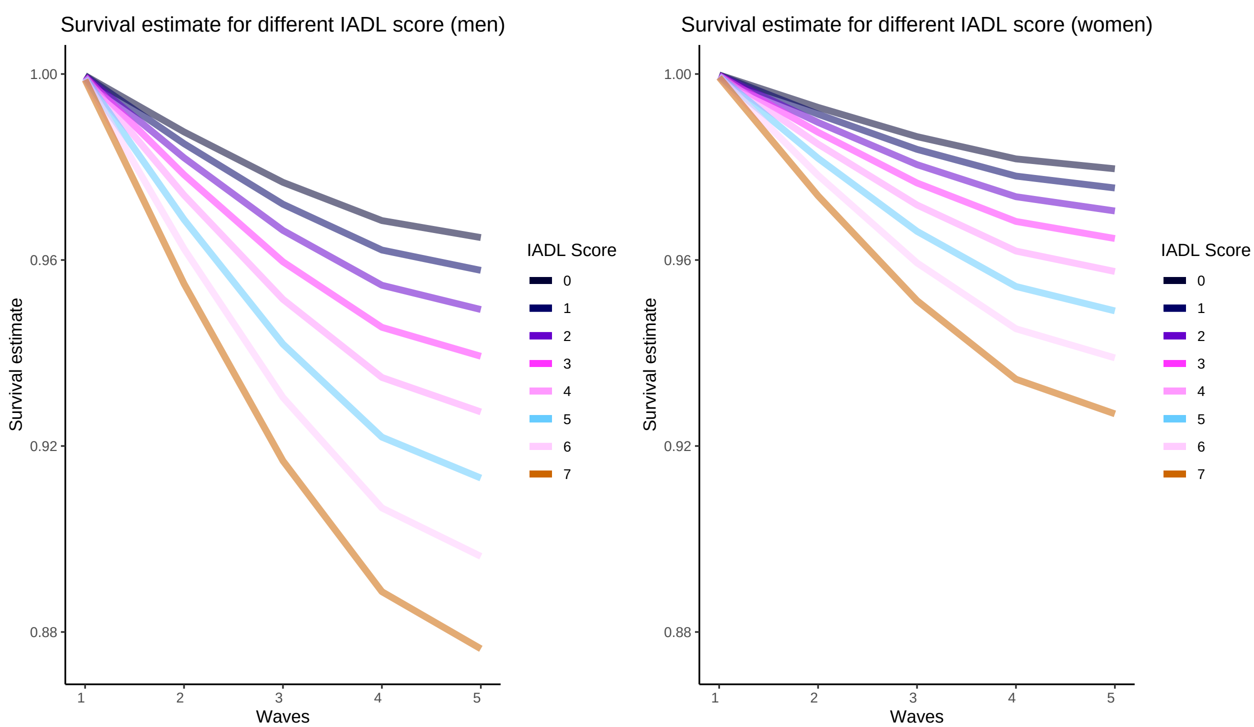


Figure 7: Survival estimate for different IADL score

physical ability of the individuals. As we can see the estimates of logistic and clog-log models are very close. According to Gumbel model, the risk of dying for men increases by about 17% compared with the risk of dying for women, which is much less than the other two models. In general. we can see that the estimates by Gumbel model are considerably smaller than the estimates by logistic and clog-log models. Among the estimates for disability and NCDs, the largest coefficient belongs to the coefficient of cancer, which increases the odds of dying among those experiencing cancer by about 3 times in both clog-log and logistic models. Contrary to our expectations, the coefficients of marital status in three categories of single, remarried and widowed, in all three models, are negative. The only category with a positive coefficient is divorced which is not statistically significant. This means that being single and widowed reduces the risk of dying compared with the reference category, i.e. being married. In fact, Grundy and Tomassini (2010) find that never-married women have lower odds of experiencing long-term illness than long-term married women. We emphasise that marriage here refers to any kind of partnership and as pointed out by Khondoker et al. (2017) being in partnership does not necessarily improve survival and the nature of the relationship is also important.

The coefficients of employment status are positive in all models and statistically significant at 5% in clog-log and logistic models, which means that the risk of dying among people who are retired, permanently sick or disabled and unemployed is higher compared with those working. The coefficients of stroke and arthritis are negative which is not in agreement with our expectations, although they are not statistically significant and therefore not significantly different from zero. We can see that AIC for all models is relatively close, although surprisingly, Gumbel has the lowest AIC.

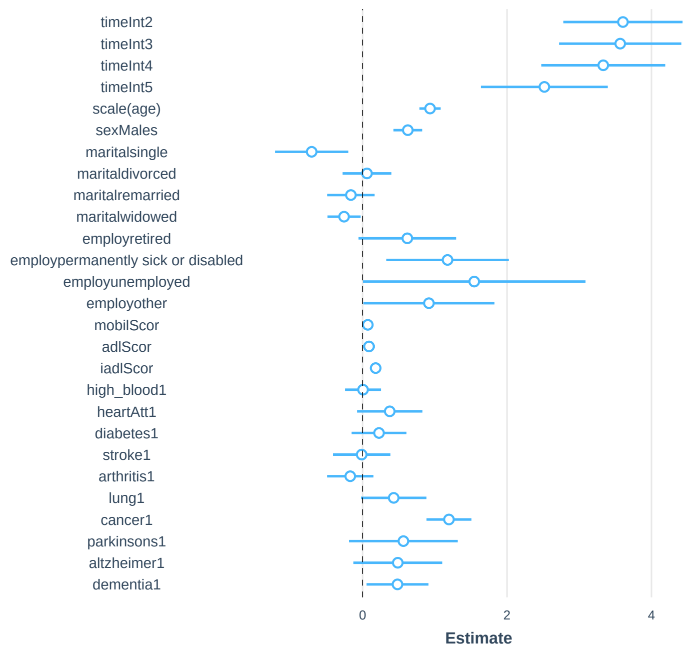
Figures 5-8 illustrate the survival curve of a hypothetical retired, married, 65-year old individual with different health conditions and physical disabilities over 5 periods. In all these figures, the projected survival curve decreases over time and the higher the disability score is, the greater the slope. In all these figures the downward slope is sharper for men than women. In Figures 5 and 6 the survival curve for men is always located below the survival curve for women regardless of mobility or ADL score. However, in Figure 7 we can observe that the survival curve for a man with IADL score of 0, 1, 2 and 3 is placed above the survival curve of a woman with IADL score of 7. By comparing the vertical axis in Figures 5-7 we can conclude that the decrease in projected survival due to IADL disability is more than the decrease in projected survival due to mobility and/or ADL disability. In Figure 8 we can see that the adverse impact of cancer on survival is much more than the impact of heart attack and chronic lung diseases both for men and women.

Mixed model with random intercept

In this section, we consider the GLMM model. Table 8 presents the coefficient estimates for person-specific random intercept model. We run this model in R, using *bglmer* to avoid the singularity problem of *glmer*. For numerical approximation we

Table 8: The estimates and confidence intervals of the coefficients of a mixed model with random intercept

Covariates	Estimate	p-value
(Intercept)	−9.7743	0.0000
Interval 2	3.6043	0.0000
Interval 3	3.5667	0.0000
Interval 4	3.3319	0.0000
Interval 5	2.5167	0.0000
Scale (Age)	0.9336	0.0000
Sex (Male)	0.6256	0.0000
Marital status:		
Single	−0.7055	0.0064
Divorced	0.0602	0.7268
Remarried	−0.1624	0.3327
Widowed	−0.2571	0.0283
Employment status:		
Retired	0.6170	0.0728
Permanently sick or disabled	1.1759	0.0066
Unemployed	1.5455	0.0493
Other	0.9178	0.0474
Disability:		
Mobility score	0.0717	0.0006
ADL score	0.0883	0.0333
IADL score	0.1805	0.0000
NCDs:		
High blood pressure	0.0058	0.9638
Heart attack	0.3754	0.1037
Diabetes	0.2269	0.2413
Stroke	−0.0129	0.9494
Arthritis	−0.1712	0.2953
Chronic lung disease	0.4304	0.0625
Cancer	1.1952	0.0000
Parkinson	0.5647	0.1416
Alzheimer	0.4868	0.1211
Dementia	0.4828	0.0273
AIC	5336.32	
Num. groups: idauniq	11984	
Variance: idauniq (Intercept)	1.149	



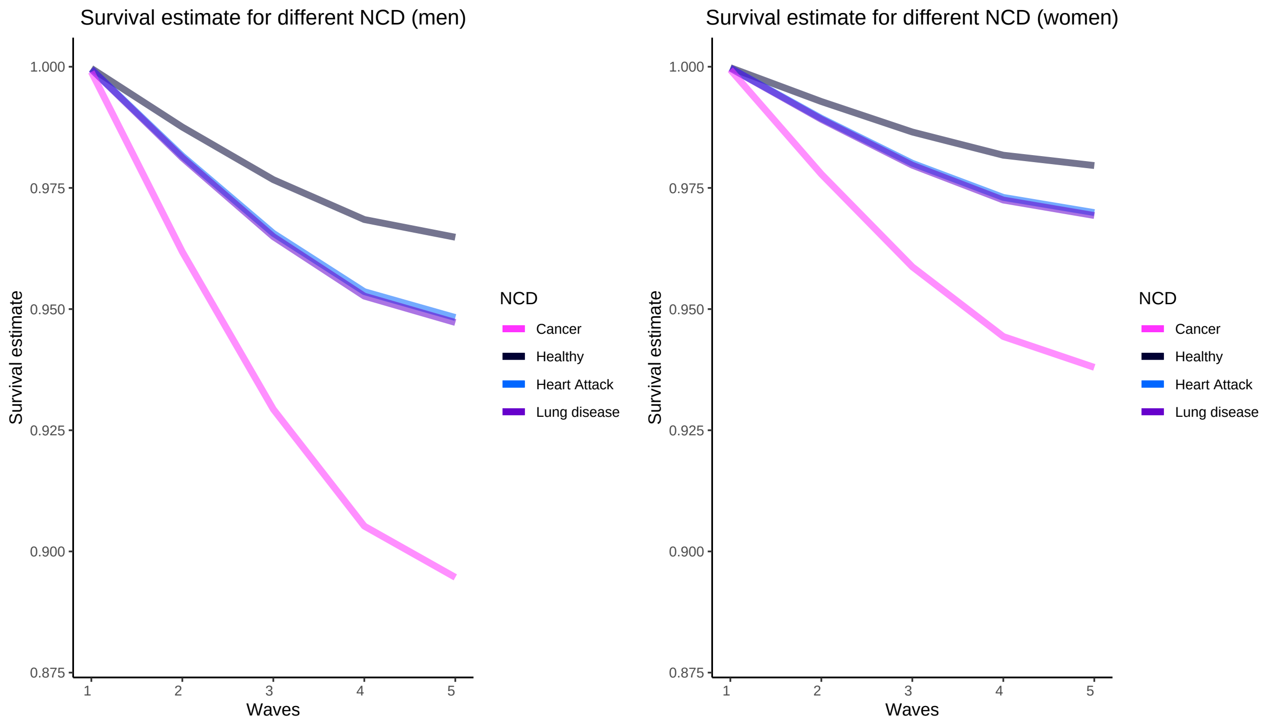


Figure 8: Survival estimate for different NCDs

use Gauss-Hermite quadrature with 20 quadrature and *bobyqa* optimisation algorithm with $200,000$ iterations. We can see a similar pattern as in Table 7. The coefficients of intervals, age, sex, disability, chronic lung diseases, cancer and dementia are positive and statistically significant with cancer having the largest coefficient among NCDs. This model is based on clog-log distribution and the AIC is very close to the AIC in Table 7. The only difference is the variance at the end of the table which implies the variability of the intercept and therefore heterogeneity among the participants. This variability can also be observed in Figure 9. In this figure, blue represents a healthy individual, pink represents an individual with IADL score of 2, light blue an individual with IADL score of 4 and light brown an individual with IADL score of 6. Further, solid lines display the survival curve of arbitrary individuals, dashed lines and double-dashed lines display survival curves for two individuals with a specific id number. If we compare Figure 9 with Figure 7 for men, we can observe that the survival projection for the average population is different from the survival projection for the individuals and therefore there are other factors that affect the projected survival, which is unique to each individual.

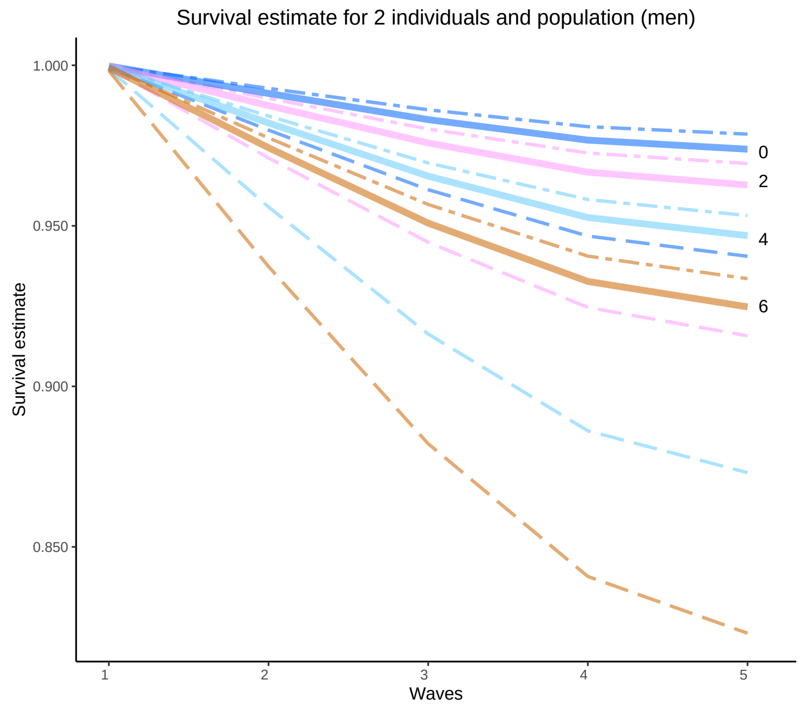


Figure 9: Dashed and double dashed lines: survival estimate for two individuals with unique id for different IADL score; solid lines: survival estimate for random individuals.

5 Conclusion

In this study, we used ELSA to look at the relationship between disabilities and NCDs with survival among men and women aged 50+. We applied discrete-time survival analysis and showed that for survey data with repeated measures, the GLMM performs better than the GLM model as GLMM takes into consideration the variability within the observations. However, AIC was slightly less than GLM models and the pattern of results was similar. The coefficients of time interval, age, sex, mobility, IADL and cancer were positive and statistically significant. We also showed that the projected survival curve for men decreases faster than the projected survival curve for women and that the decrease in survival due to IADL disability and cancer is much more than other causes.

References

[1] Austin, P.C. (2017) A tutorial on multilevel survival analysis: methods, models and applications. *International Statistical Review*, 85, 2, 185–203.

[2] Blake, M., Bridges, S., Hussey, D. and Mandalia, D. (2015) The dynamics of ageing: The 2010 English Longitudinal Study of Ageing (wave 5). *London: NatCen*

Social Research.

[3] Buttery, A.K., Du, Y., Busch, M.A., Fuchs, J., Gaertner, B., Knopf, H. and Scheidt-Nave, C. (2016) Changes in physical functioning among men and women aged 50-79 years in Germany: an analysis of National Health Interview and Examination Surveys, 1997–1999 and 2008–2011. *BMC Geriatrics*, 16(1): 205.

[4] Demakakos, P., Biddulph, J.P., Bobak, M. and Marmot, M.G. (2015) Wealth and mortality at older ages: a prospective cohort study. *J Epidemiol Community Health*, 70(4): 346–353.

[5] Grundy, E.M.D. and Tomassini, C. (2010) Marital history, health and mortality among older men and women in England and Wales. *BMC Public Health*, 10: 554.

[6] Kessler, M., Thumé, E., Scholes, S., Marmot, M., Facchini, L.A., Nunes, B.P., Machado, K.P., Soares, M.U. and de Oliveira, C. (2020) Modifiable risk factors for 9-year mortality in older English and Brazilian adults: The ELSA and SIGa-Bagé ageing cohorts. *Scientific reports*, 10(1): 1–13.

[7] Khondoker, M., Rafnsson, S.B., Morris, S., Orrel, M. and Steptoe, A. (2017) Positive and negative experiences of social support and risk of dementia in later life: An investigation using the English Longitudinal Study of Ageing. *Journal of Alzheimer's Disease*, 58 (2017): 99–108.

[8] Lin, S.F., Beck, A.N., Finch, B.K., Hummer, R.A. and Master, R.K. (2012) Trends in US older adult disability: exploring age, period, and cohort effects. *American Journal of Public Health*, 102(11): 2157–2163.

[9] Pongiglione, B., De Stavola, B.L., Kuper, H. and Ploubidis, G.B. (2016) Disability and all-cause morality in the older population: evidence from the English Longitudinal Study of Ageing. *Eur J Epidemiol*, (2016) 31: 735–746.

[10] Pongiglione, B., Ploubidis, G. and De Stavola, B. (2017a) Disability-free life expectancy between 2002 and 2012 in England: trends differ across genders and levels of disability. In *2017 International Population Conference*. IUSSP.

[11] Pongiglione, B., Ploubidis, G. and De Stavola, B. (2017b) Levels of disability in the older population of England: Comparing binary and ordinal classifications. *Disability and Health Journal*, 10 (2017): 509–517.

[12] Rafnsson, S.B., Orrell, M., d'Oris, E., Hogervorst, E. and Steptoe, A. (2020) Loneliness, social integration, and incident dementia over 6 years: Prospective findings from the English Longitudinal Study of Ageing. *Journals of Gerontology: Social Sciences*, 27(1): 114–124.

[13] Rogers, N.T., Demakakos, P., Taylor, M.S., Steptoe, A., Hamer, M. and Shankar, A. (2016) Volunteering is associated with increased survival in able-bodied participants of the English Longitudinal Study of Ageing. *J Epidemiol Community Health*, 70(6): 583–588.

[14] Steptoe, A., Breeze, E., Banks, J. and Nazroo, J. (2013) Cohort profile: The English Longitudinal Study of Ageing. *International Journal of Epidemiology*, 42: 1640–1648.

[15] Steptoe, A., Deaton, A., Stone, A.A. (2015) Psychological wellbeing, health and ageing. *Lancet*, 385(9968): 640.

[16] Tutz, G. and Schmid, M. (2016) Modeling discrete time-to-event data, Springer series in Statistics,