Survival of the Fittest?

An empirical investigation into the determinants of disability and mortality in Great Britain

Martin Karlsson[1], Les Mayhew, Ben Rickayzen

June 30th 2006

Abstract

The main aim of this paper is to identify factors - socioeconomic and others - that determine the dynamics of disability and mortality amongst older people in the United Kingdom. Furthermore, we analyse the extent to which disability and mortality are related so that estimates based only on survivors produce biased estimates.

We define disability as failing one or more ADLs, and estimate determinants of survival and disability jointly. In order to account for the possibility of correlation in unobservables between the two equations, we make use of a Heckman bivariate probit model, where the correlation between the two error terms is a paramater to be estimated. Our analysis is based on three different waves of the British Household Panel Survey (BHPS).

We find that the most important determinants of both survival and disability are smoking and previous disability. Furthermore, we cannot reject the hypothesis of independence between the two variables in the short term, whereas the correlation is extremely strong in the long term. This result suggests that estimates based only on survivors will underestimate the impact of the various risk factors.

Keywords: disability, mortality, risk factors

JEL Classificartion: I11, H51

[1] Corresponding address: Martin Karlsson, Faculty of Actuarial Science and Insurance, Cass Business School, City University, 106 Bunhill Row, London EC1Y 8TZ, UK. Phone number: +44 (0)207 040 5234, Fax Number +44 (0)20 7040 8572. Email: p.m.karlsson@city.ac.uk

1 Introduction

Most developed countries' populations are ageing rapidly with consequent implications for public spending on long term care (LTC), pensions and health care. The UK dependency ratio (the number of retired people per 100 people of working age) is projected to increase from 24 today to 38 in 2040. Although substantial, the increase is lower than in many other countries. In Japan, for instance, the ratio is projected to increase from 30 today to 65 in 2040 (United Nations, 2002).

Such demographic changes are expected to have a significant impact on the demand for LTC. Most consumers of LTC are over age 80; for example, in England, almost 80 per cent of care home inhabitants belong to this age group (Bajekal, 2002). Since increasing life expectancy causes this group to grow at a faster rate than the general retired population, there is concern that the demographic burden could make the current system of financing LTC unsustainable. Indeed, in the UK, there has been a recent trend towards concentrating resources only on individuals with severe disability (Karlsson et al, 2004).

Still, relatively little is known about long term trends and the determinants of the disablement process. One important issue that has not yet been resolved is the long term trends in healthy life expectancy and disabled life expectancy. Three competing hypotheses have been proposed. The most optimistic one, suggesting a compression of morbidity, is due to Fries (1980). According to this perspective, adult life expectancy is approaching its biological limit so that if disability spells can be postponed to higher ages the result will be an overall reduction in the time spent disabled. By contrast, Gruenberg (1977) suggested an expansion of morbidity based on the argument that the observed decline in mortality was mainly due to falling accident rates. The third hypothesis was proposed by Manton (1987) according to whom the development in mortality and morbidity is a combination of the two, which could lead to an expansion of the time spent in good health as well as the time spent in disability.

Official statistics, however, are surprisingly inconclusive as to which of the three hypotheses prevails in reality (Bone et al, 1995, Bebbington & Darton, 1996, Bebbington and Comas-Herrera, 2000). In general, results seem to be sensitive to the definition of disability (ADLs or limiting long-standing illness) as well as to the severity of disability taken into account. Despite this ambiguity in the statistics, the long term trends have very strong implications for the future funding of long term care. In a long-term projection model, Karlsson et al (2006) find that an optimistic scenario ('compression of morbidity') imply some 2 million disabled older people fewer than the most pessimistic scenario ('expansion of morbidity'). The implications for public finances are similar: in the pessimistic scenario, tax rates necessary to finance the current system of LTC provision would have to increase by around 80 per cent, whereas virtually no increase would be necessary in the optimistic scenario.

The projections of Karlsson et al (2006) are based on the Rickayzen & Walsh (2002) model, which is one of the few models available for projections of future LTC costs in the UK. The model has two limitations however, which the current paper attempts to deal with. Firstly, it is based on a very old dataset – the Office for Population Censuses and Surveys (OPCS) survey from the 1980s. Secondly, it assumes individual within a certain age and gender group are homogenous. Hence, it does not lend itself to studies of subgroups of the population, or to analyse the implications of changes in various risk factors.

In this paper, we make use of three waves of the British Household Panel Survey (BHPS) data in order to study the determinants of disability and mortality over time. Our research has three main aims. Firstly, we want to analyse the extent to which disability and mortality are correlated in a way such that estimates based only on survivors become biased. Secondly, we want to analyse the extent to which typical risk factors have similar effects on males and females. Thirdly, we wish to analyse whether these risk factors affect disabled and non-disabled people differently.

Our methodological approach is to use Heckman's (1979) model applied to a context of limited dependent variables. Hence, we estimate a censored bivariate probit model, where censoring is based on survival (i.e. only disability status of survivors is observed). The correlation in unobservables between the two equations (survival and disability) is a parameter to be estimated.

In many rigorous studies of dynamics of disability, the problem of "selective mortality" – i.e. that survivors are not representative of the entire population – has been ignored (e.g. Contoyannis et al, 2004 and Börsch-Supan et al, 1992). Papers that do account for mortality generally tend to incorporate death as a separate disability category – being the most severe category (e.g. Diehr et al, 2003, Karlsson et al, 2006b). This way the problem of selective mortality is overcome, but the approach is not without problems since it constrains coefficients with respect mortality to be the same as the coefficients with respect to disability. In this paper, instead, we treat survival as a selection function which is potentially correlated with the disability function. This way we get unbiased estimates of the determinants of disability and mortality, as well as an estimate of the correlation coefficient between the two.

Concerning the two other issues, there are numerous studies which have analysed gender differences in the disablement process. Generally, there are differences between men and women in their exposure to risk factors. Furthermore, some studies find that the risk factors have systematically different effects on males and females. Consequently, Strawbridge et al (1993) found that education and income were associated with disability for men but not for women. On the other hand, Mor et al (1989) found that education was related to functional decline for women but not men and Unger et al (1999) found that poor social networks was a stronger predictor of disability for men than for women. Hence, the empirical evidence is inconclusive and our paper attempts to fill this gap.

Our main findings can be summarised as follows. We find that selective mortality seems not to be a problem in the short term, whereas the correlation between survival and disability is extremely strong over a 10 year horizon. Furthermore, we are not able to reject the null hypothesis of coefficient equality for men and women in the short term, whereas parameter estimates differ systematically in the long term. Finally, regarding the initial disability state, we cannot reject the hypothesis that risk factors have the same impact on people initially healthy as on those initially disabled.

The rest of the paper is structured as follows. In the next section, we present out econometric model. In Section 3, we give a brief overview of the dataset and some descriptive statistics. Section 4 provides the main results and some hypothesis testing. Section 6 concludes and points out directions for future research.

2 Estimation Method

2.1 The Estimator

We want to explain how disability at a future stage in life is explained by current personal characteristics, while at the same time recognising that survivors are not randomly chosen. Our dependent variable - disability - is a binary variable which is censored for individuals who do not survive until the later date. This calls for estimating the selection mechanism as well, since otherwise parameter estimates of determinants of disability would be biased.

Hence, for the disability dimension, we use a simple probit model with the latent continuous variable d^*:

$$d^* = \alpha_d + X_d' \beta_d + \varepsilon_d \tag{1}$$

where α_d is a constant, X_d' is a matrix of demographic and personal characteristics and β_d is a vector of coefficients. In a similar manner, the survival probability is modelled as:

$$s^* = \alpha_s + X_s' \beta_s + \varepsilon_s \tag{2}$$

The error terms are assumed to have a bivariate normal distribution with correlation coefficient ρ:

$$\begin{bmatrix} \varepsilon_d \\ \varepsilon_s \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right) \tag{3}$$

Observed outcomes are determined according to

$$d = \begin{cases} 1 & \text{if } d^* > 0 \\ 0 & \text{if } d^* \leq 0 \end{cases} \tag{4}$$

$$s = \begin{cases} 1 & \text{if } s^* > 0 \\ 0 & \text{if } s^* \leq 0 \end{cases} \tag{5}$$

Furthermore, we have a censored probit model so that d is observed if and only if $s = 1$. The likelihood function of this system of equations reads:

$$\ln L(\beta_d, \beta_s, \rho) = \sum_{i \in I} \left\{ \begin{array}{c} s_i d_i \ln F(X_{di}\beta_d, X_{si}\beta_s, \rho) \\ +s_i(1-d_i)\ln[\Phi(X_{si}\beta_s) - F(X_{di}\beta_d, X_{si}\beta_s, \rho)] \\ +(1-s_i)\ln\Phi(-X_{si}\beta_s) \end{array} \right\} \tag{6}$$

where $F(\cdot)$ is the cumulative distribution function of the bivariate normal distribution, $\Phi(\cdot)$ is the cumulative distribution function of a univariate normal distribution and I is the set of individuals belonging to the sample.

There are several advantages to this modelling approach, which become clear if we compare it with alternative approaches. As a first alternative, the model could have been estimated portraying health and disability as one dimension, in which case an ordered probit model would have been appropriate. With this approach, the interdependence between disability and death is allowed for, but it is assumed that the coefficients are the same for disability and death. If, on the other hand, the two equations had been estimated separately, we would be ignoring the fact that survivors are a non-random sample, and parameter estimates would have been biased.

Since we are using a panel dataset, the panel structure could also have been exploited by using panel data methods and estimating a fixed or random effects model. Given that we want to model self-selection, however, this would require a very complex model. Instead, we condition on previous disability - which should, to some extent, control for unobserved heterogeneity.

The censored bivariate probit model we have outlined above is not entirely straightforward to use for prediction purposes, since the bivariate truncated normal distribution does not have an analytical solution. It is, however, possible to simulate a bivariate normal distribution and then use the lower Choleski decomposition of the covariance matrix in (3) to simulate probabilities. This way approximate probabilities for an individual being in a certain state, conditional upon individual characteristics X_d' and X_s', can be calculated.

2.2 Hypothesis Testing

We use a sample consisting of permanent members of the BHPS who were over 50 years old in the first wave (1991). It is unclear, however, the extent to which these individuals can be pooled in one single regression. This is of particular relevance for the gender dimension (i.e. risk factors and selection might work differently on men and women) and in the disability dimension: individuals who are disabled at the outset might follow other dynamics than those who are initially healthy. Hence, we performed likelihood ratio tests to determine whether pooling should allowed in these two dimensions. The test statistic

$$LR = 2\left(\ln L_1 - \ln L_2\right) \tag{7}$$

where L_1 is the model estimated with separate coefficient estimates for males (or alternatively, those initially disabled) and L_2 is the baseline model where gender-specific coefficients are constrained to be equal to zero. Under weak regularity conditions, the Likelihood Ratio test statistic is approximately chi-square distributed with degrees of freedom equal to the difference between the dimensions of the unrestricted and restricted models (i.e., the number of gender- or disability-specific parameters).

3 The Dataset

3.1 Variables Included

We use three separate waves of the BHPS; 1991, 1996 and 2001. All independent variables are taken from the first wave, whereas we have recorded survival and disability for the two subsequent waves. The variables used for estimation are presented in *Table 1*. Since identification in the Heckman probit model requires that a different set of variables are used in the selection equation (2), we added squared age in this equation.

Table 1. Definition of Variables.

Variable	Explanation
s	Alive at $t + x$
d	At least one ADL failure at $t + x$
AGE	1991 minus birthyear
GENDER	Male or female
EDUCATION	Reached at least GCSE level
SMOKER	Smokes cigarettes regularly
INITIAL STATE	Value of d in 1991
HOME OWNER	Home ownership in 1991
COHABITATION	Individual married or cohabiting in 1991.
REGION	Region dummies for Scotland, Wales, North England, South England and London

The sample has a complete record of deaths, but for survivors, missing values are sometimes a problem for the disability variable. We used the following approach: if the disability variable had a missing value in 1996 or 2001, and the individual was not dead, the most recent observation of disability status available was used. This way virtually all gaps in the data could be filled in.

3.2 Descriptive Statistics

In the following, we will provide some summary statistics of the panel. We start out by showing the main variable of interest - disability - and how it evolves with age in *Table 2*.[2] Just as in the estimates, we have defined disability quite generously here as failing one or more ADLs.

Table 2. Health Status (ADLs) by age, 1991.

Age	*Healthy*	*Disabled*	*Total*
<60	841	188	1,029
	81.73	*18.27*	*100.00*
60-69	788	234	1,022
	77.10	*22.90*	*100.00*
70-79	529	247	776
	68.17	*31.83*	*100.00*
80-89	171	132	303
	56.44	*43.56*	*100.00*
90+	13	16	29
	44.83	*55.17*	*100.00*
Total	**2,342**	**817**	**3,159**
	74.14	*25.86*	*100.00*

Table 2 shows the well documented relationship between health and age. For instance, among people in their fifties, the prevalence of physical impairments is less than 20 per cent and at the highest ages the majority of people have at least one impairment.

Next, we look at the role of cohabitation. In Table 3, we cross-tabulate the initial wave by health status and cohabitation status; again, the disabled status is assumed to be when one or more ADLs are failed. The two seem to be correlated; a person not cohabiting is fifty per cent more likely to be disabled than a person who is cohabiting.

Furthermore, the initial cohabitation state seems to be a reasonable predictor of the health

[2] In principle, the data could also be partitioned by gender. Due to the few observations in some age brackets, however, it is better to present the pooled dataset. Males and females exhibit the same patterns of gradually deteriorating health, but female survivors tend to have slightly worse health than males.

Table 3. Health Status (ADLs) by cohabitation status, 1991.

cohabit	*Healthy*	*Disabled*	*Total*
No	762	372	1,134
	67.20	*32.80*	*100.00*
Yes	1,580	445	2,025
	78.02	*21.98*	*100.00*
Total	2,342	817	3,159
	74.14	*25.86*	*100.00*

status in subsequent years, including death. In *Table 4* we cross-tabulate the cohabitation status in 1991 with the health status in 1996. Clearly, people who were cohabiting in 1991 had a higher chance of being alive and healthy in 1996. The mortality rate in particular seems to be high for non-cohabiting people when compared to cohabiting people.

Finally, we look at the relationship between health and education. Figures are presented in *Table 5*. The Education variable reflects the self reported educational attainment, where for simplicity we have merged the educated categories into one single group, corresponding to an educational level of at least GCSEs (or the predecessors: GCE, Higher School certificate etc). The health variable is disability in 1996. As expected, a higher educational attainment is correlated with better health. The effect of education seems to be particularly strong for moderate disability, where the prevalence amongst non-educated people is more than twice as high as the corresponding figure for educated people.

Table 4. Health Status 1996 (ADLs) by cohabitation status 1991.

Cohabit	*Healthy*	*Moderate*	*Severe*	*Dead*	*Total*
No	739	60	194	141	1134
	65.2%	*5.3%*	*17.1%*	*12.4%*	*100.0%*
Yes	1,539	79	256	151	2,025
	76.0%	*3.9%*	*12.6%*	*7.5%*	*100.0%*
Total	**2,278**	**139**	**450**	**292**	**3,159**
	72.1%	*4.4%*	*14.2%*	*9.2%*	*100.0%*

Table 5. Health Status 1996 (ADLs) by educational attainment (GCSE+ equivalent) 1991.

Education	*Healthy*	*Moderate*	*Severe*	*Dead*	*Total*
No	1443	113	338	227	2,121
	68.0%	*5.3%*	*15.9%*	*10.7%*	*100.0%*
Yes	835	26	112	65	1,038
	80.4%	*2.5%*	*10.8%*	*6.3%*	*100.0%*
Total	**2,278**	**139**	**450**	**292**	**3,159**
	72.1%	*4.4%*	*14.2%*	*9.2%*	*100.0%*

4 Results

In this section we report the estimation results and discuss hypothesis testing. Our primary results are set out in *Table 6*[3]. They show the determinants of disability and survival for the entire sample, over five and ten year horizons.

Table 6. Estimation Results.

	T+5		**T+10**	
Variable	*Survival*	*Disability*	*Survival*	*Disability*
Constant	3.885	-1.471	2.963	-2.689
	2.67**	-4.90**	2.83**	-12.17**
Male	-0.3136	-0.0105	-0.3401	0.1334
	-4.33**	-0.17	-5.96**	2.62**
Cohabitation	-0.0151	-0.1369	0.1246	-0.0403
	-0.19	-2.11*	2.03*	-0.71
Smoker	-0.2925	0.2045	-0.4103	0.3214
	-3.63**	2.88**	-6.39**	5.46**
Disabled0	-0.5710	1.610	-0.5353	1.2358
	-8.08**	12.20**	-9.17**	19.61**
Owner	0.1887	-0.0520	0.1175	-0.1146
	2.50*	-0.77	1.96	-2.03*
Education	-0.0050	-0.1453	0.0376	-0.1830
	-0.06	-2.29*	0.61	-3.38**
Age	-0.0312	0.0109	-0.0049	0.0399
	-0.74	2.42*	-0.16	14.31**
Age2	-0.0064		-0.0381	
	-0.21		-1.73	
South England	0.3000	-0.2598	0.2626	-0.3201
	2.51*	-2.43*	2.72**	-3.59**
North England	0.1992	-0.1019	0.0758	-0.1333
	1.79	-1.02	0.83	-1.55
Scotland	0.1920	0.0780	0.0049	-0.0762
	1.13	0.54	0.04	-0.59
Wales	0.2007	-0.1409	0.1323	-0.1624
	1.32	-1.05	1.07	-1.42
Athrho	0.4152		-6.216	
	1.01		-0.01	
N	3,159		3,159	
LR-Chi2	1.19		8.65**	

In general, the parameter estimates in *Table 6* have the expected signs. As expected, males have a much lower survival probability than females, both in the short and the long term. Con-

[3] *Disabled0* is initial disability status. London is the reference category. Two asterisks next to a t value represents significance at the one per cent level; one asterisk represents significance at the five per cent level.

cerning disability, however, males seem to have worse prospects only in the long term. Cohabitation reduces the probability of becoming disabled in the short term, and increases the survival probability in the long term. The effects are, however, weak in comparison to smoking, which has a substantial negative effect on survival probabilities and health over both time horizons. In the survival dimension, the effects of smoking are comparable in magnitude to the difference between men and women in survival probability.

As expected, previous disability has a strong negative impact in both dimensions. Home ownership is connected with an increase in survival prospects in the long term, and with a reduction in disability probability in the long term. Education, on the other hand, seems to only reduce the risk of disability without having an effect on survival. This is an unusual finding which requires further investigation.

Regarding regional effects, in South England the health features are significantly different from the rest of the country. People in this area enjoy much lower mortality and disability rates than the rest of the population.

Finally, we perform a likelihood ratio test to establish whether the two estimating equations are independent. The null hypothesis of no independence cannot be rejected for the short term, whereas it is rejected at a high level of confidence for the long term. The estimated value of ρ is -0.9999428; hence, even after controlling for the independent variables, there is strong selection into survival based on unobserved factors. Accordingly, an econometric model that does not allow for this self-selection would seriously underestimate the effects of the risk factors.

4.1 Hypothesis Testing

The assumption that pooled estimates can be used clearly requires justification. It would be expected that males are systematically different from females and that people already disabled are systematically different from those who are not. In order to test these assumptions, the likelihood ratio test outlined in *Section 2.2* was used. We estimated two models, one with the restriction that coefficients are the same for males and females, and one without that restriction. We found that none of the gender-specific coefficients were significant in any of the estimating equations - and this goes for the five year horizon as well as the ten year one. Results from the likelihood ratio test are provided in *Table 7*.

Table 7. Test Result. LR Test for pooling of males and females.

	T+5	**T+10**
Likelihood, full model	-2095.15	-2672.29
Likelihood, restricted model	-2102.23	-2683.55
LR Chi2(df)	14.15(13)	22.52(13)
P	0.3633	0.0478

Interestingly, in the short term, we cannot reject the null hypothesis that coefficients for males and females are equal; but in the ten year perspective, we reject the equality assumption at a five per cent level of confidence. Hence, for the long term horizon, separate estimations should be made for males and females. We return to this issue in *Section 4.2* below.

In a similar vein, it seems plausible to assume that people initially disabled have different parameter values than people who were healthy at the outset. In order to allow for this possibility, we estimated the model with different coefficients for people initially disabled. Again, the disability-specific parameters were generally insignificant, with the only exception that in the five year perspective, home ownership seems to increase the survival probabilities for disabled people more than for non-disabled people. We present the results from the likelihood ratio test in *Table 8*.

Table 8. Test Result. LR Test for pooling of disabled with non-disabled.

	T+5	T+10
Likelihood, full model	-2094.35	-2676.91
Likelihood, restricted model	-2102.23	-2683.55
LR Chi2(df)	15.75(13)	13.29(13)
P	0.2929	0.4257

In conclusion, we have found that coefficients do not differ significantly between different subsamples, with the sole exception of males and females for the ten year horizon.

4.2 Separate estimates for males and females

According to the test results set out in *Table 7* above, estimates should be made separately for males and females for the ten year horizon. We present the results in *Table 9*.

When males and females are dealt with separately, we notice several differences compared to the original estimates in *Table 6*. Firstly, the cohabitation variable loses significance. Whereas cohabitation is connected with significantly better survival probabilities in the pooled sample, there is no significant effect on disability or survival in the gender-specific regressions. Given the striking gender differences in cohabitation among older people, this result suggests that the cohabitation parameter was previously picking up a gender effect. On the other hand, the parameter estimates have the expected signs, with the notable exception of male disability where the effect of cohabitation is estimated to be positive (but insignificant).

As regards smoking, we find that the effects are quite different for the two sexes. The results are still strongly significant for disability as well as survival. For males, however, the impact on survival prospects seems to be much stronger than the effect on disability, whereas coefficients for females are similar in the two dimensions. This result suggests that public health interventions aimed at reducing smoking could have quite different effects on males and

Table 9. Estimation Results, Males and Females separately.

	Males		**Females**	
Variable	*Survival*	*Disability*	*Survival*	*Disability*
Constant	3.332	-2.174	1.554	-2.604
	1.72	-3.48**	1.11	-8.63**
Cohabitation	0.1524	0.1620	0.0980	-0.1191
	1.59	1.25	1.20	-1.63
Smoker	-0.4902	0.2705	-0.3558	0.3204
	-5.15**	2.20*	-3.97**	3.94**
Disabled0	-0.6094	1.3458	-0.4945	1.2545
	-6.74**	11.39**	-6.34**	15.05**
Owner	0.1067	-0.0121	0.1071	-0.1885
	1.13	-0.12	1.35	-2.53*
Education	0.0808	-0.2625	-0.0110	-0.1442
	0.88	-2.93**	-0.12	-1.91
Age	-0.0192	0.0323	0.0314	0.0366
	-0.33	2.56*	0.77	9.63**
Age2	-0.0353		-0.0594	
	-0.83		-2.04*	
South England	0.3991	-0.5434	0.2043	-0.1178
	2.72**	-3.55**	1.52	-0.98
North England	0.2917	-0.4720	-0.0485	0.1469
	2.09*	-3.21**	-0.39	1.27
Scotland	-0.1060	-0.2746	0.1621	0.0695
	-0.53	-1.20	0.83	0.40
Wales	0.2991	-0.4382	0.0381	0.0562
	1.56	-2.20*	0.23	0.37
Athrho	-0.6955		-4.965	
	-1.09		-0.09	
N	1,401		1,758	
LR-Chi2	2.19		5.30*	

females.

Turning to home ownership, we find that it only has a significant effect on female disability in the long term, whereas the effect is negligible for males. Education, on the other hand, seems to have a significant impact on males only.

Finally turning to the regional dummies, we find some interesting patterns in the gender-specific results. Firstly, It is now not only South England that is significantly different from London. Now we observe that males in North England also have significantly better health and survival probabilities than their counterparts in London. Welsh males have significantly lower probabilities of becoming disabled than London males.

On the other hand, for females, the patterns are quite different. For people in South England, the estimates have the same sign as in the original estimation, but the overall significance is now lost; and no other region has significantly different health either. This result suggests that regional variation in health and mortality is predominantly a male phenomenon. One possible explanation to this gender discrepancy is that there is more regional variation in working conditions for males than for females.

Concerning the interdependence between the two estimating equations, we find that there is very strong correlation in the residuals for females - the parameter is estimated as $\rho = -0.999929$. For males, on the other hand, we cannot reject the hypothesis of independence between the two equations. In conclusion, estimates based only on survivors would severely bias female coefficients but not male ones.

4.3 Illustration

We have found that estimating the effects of various risk factors only using survivors is problematic, since - at least in the long term - disability and survival are correlated events, even after controlling for observable characteristics. The next question is how important this problem is. In order to give a crude estimate of the importance of the correlation, we compare estimates with and without the correlation assumed in *Figure 1*. The two bars to the left show predicted probabilities of being in various categories when independence between survival and disability is assumed. The two bars to the right show the same predictions in the model with selective mortality. We assume a single 60 year old woman with no secondary education who is not cohabiting at the outset, as our example. The bars in *Figure 1* show her predicted probabilities of being in one of the three categories ten years later.

Figure 1 shows that, by ignoring the selection effects, a simple probit estimation seriously

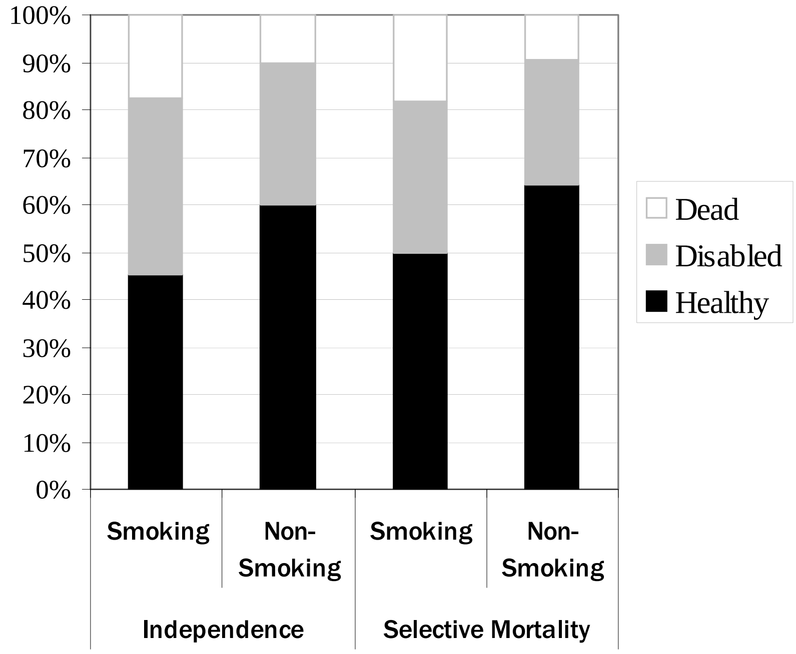

Figure 1. Effects of smoking with and without selective mortality.

overestimates the reduction in disability that can be brought about by reducing smoking. The simple probit model predicts that eliminating smoking would bring about a 12 percentage point reduction in the prevalence of disability among survivors, whereas the corresponding figure for the Heckman approach is a reduction by nine percentage points only. Similar discrepancies appear when other combinations of the independent variables are used. Accordingly, the biases introduced by assuming independence are not enormous but are significant.

5 Conclusion

The availability of longitudinal datasets allows us to gain a better understanding of the processes that determine the deterioration of physical function at older ages. In this paper we have made use of the BHPS in order to analyse how the disablement process relates to mortality risk. Furthermore, we have analysed the extent to which typical risk factors have a different impact on males and females, or on people initially healthy as compared to people who have functional limitations at the outset.

Our findings are mixed. We find that there is a potential correlation in observables between disability and survival - at least in the long term - which calls into question previous studies that have been based only on survivors. Furthermore, we are not able to reject the hypothesis that risk factors have the same impact on disabled and non-disabled people - a finding which, of course, simplifies the analysis of the dynamics of disability great deal. Concerning the gender dimension, on the other hand, we find that females and males suffer differently from different risk factors in the long term. For instance, home ownership seems to be more beneficial to females than to males, whereas education reduces disability risk for males only. Also regional differences in disability and mortality seem to be more pronounced for males than for females.

The model presented here suffers from two important limitations. Firstly, it is static and hence cannot be used without adjustment to produce reliable simulations of the entire population. Secondly, even if we have allowed for selection effects in the survival function, we have not allowed for self selection into the independent variables. Hence, we cannot distinguish whether the effects we observe (i.e. concerning risk factors such as smoking or housing tenure) reflect causality or only correlation. These issues could be addressed by future research.

6 References

BAJEKAL, M. (2002), *Care Homes and their Residents*, London: The Stationery Office.

BEBBINGTON, A. C. & DARTON, R. A. (1996). *Healthy life expectancy in England and Wales: recent evidence*. PSSRU Discussion Paper 1205.

BEBBINGTON, A & COMAS-HERRERA, A. (2000). *Healthy Life Expectancy: Trends to 1998, and the Implications for Long Term Care Costs*, PSSRU Discussion Paper 1695 (London School of Economics).

BONE, M., BEBBINGTON, A. C., JAGGER, C., MORGAN, K. & NICOLAAS, G. (1995). *Health expectancy and its uses*. London: HMSO.

BOERSCH-SUPAN, A., HAJIVASSILIOU, V., KOTLIKOFF, L. AND MORRIS, J. (1992). “Health, Children, and Elderly Living Arrangements: A Multiperiod-Multinomial Probit Model with Unobserved Heterogeneity and Autocorrelated Errors.” in *Topics in the Economics of Aging*. (ed.) David Wise. Chicago: University of Chicago Press.

CONTOYANNIS, P., JONES, A.M. & RICE, N. (2004), ”Simulation-based inference in dynamic panel probit models: An application to health”, *Empirical Economics* 29, 49–77

DIEHR, P. & PATRICK, D. L. (2003), ”Trajectories of Health of Older Adults Over Time: Accounting fully for death”, *Annals of Internal Medicine* 139, 416-20.

FRIES J (1980), “Aging, Natural Death and the Compression of Morbidity.” *New England Journal of Medicine* 303:130-5.

GRUENBERG EM (1977). ”The failures of success”. *Milbank Memorial Foundation Quarterly*/Health and Society, vol.55, pp.3-24.

HECKMAN, J. (1979), ”Sample selection bias as a specification error”, *Econometrica* 47, 153-61.

KARLSSON M., MAYHEW, L., PLUMB, R. & RICKAYZEN, B. (2004), *An International Comparison of Long-Term Care Arrangements: An Investigation into the Equity, Efficiency and Sustainability of the Long-Term Care Systems in Germany, Japan, Sweden, the United Kingdom and the United States*. Working Paper, CASS Business School, City University.

KARLSSON M., MAYHEW, L. & RICKAYZEN, B. (2006), ”Future costs for long-term care. Cost projections for long-term care for older people in the United Kingdom”, *Health Policy* 75, 187–213

KARLSSON M., MAYHEW, L. & RICKAYZEN, B. (2006b), ”In Sickness and in Health? Dynamics of Health and Cohabitation in the United Kingdom”, mimeo, Cass Business School.

MANTON KG (1987). ”Response to ’an introduction to the compression of morbidity”’. *Gerontologica Perspecta*, vol.1, pp.23-30.

MOR V, MURPHY J, MASTERSON-ALLEN S, WILLEY C, RAZMPOUR A, JACKSON ME, GREER D, KATZ S. (1989), ”Risk of functional decline among well elders”, *J Clin Epidemiol* 1989;42:895-904.

RICKAYZEN, B. D. & WALSH, D. (2002), “A Multi-State Model of Disability for the U.K.: Implications for Need for Long Term Care for the Elderly”, *British Actuarial Journal*, 8, II.

STRAWBRIDGE WJ, KAPLAN GA, CAMACHO TC, COHEN RD (1992), ”The dynamics of disability and functional change in an elderly cohort: Results from the Alameda County Study”. *J Am Geriatr Soc* 1992;40:799-806.

UNGER JB, MCAVAY G, BRUCE ML, BERKMAN L, SEEMAN T (1999), ”Variation in the impact of social network characteristics on physical functioning in elderly persons: MacArthur Studies of Successful Aging”, *J Gerontol Soc Sci* 54B:S245-51.

UNITED NATIONS (2002), Population Database, http://esa.un.org/unpp/