HEDG
HEALTH, ECONOMETRICS AND DATA GROUP

THE UNIVERSITY of York

WP 11/30

Survival expectations, subjective health and smoking: evidence from European countries

Silvia Balia

September 2011

york.ac.uk/res/herc/hedgwp

Survival expectations, subjective health and smoking: evidence from European countries.

Silvia Balia*
Università di Cagliari

September, 2011

1 Abstract

This work aims to assess risk perception of smokers in reporting survival expectations and subjective health. In particular, the analysis investigates individuals' perception of smoking effects in the short and long-term and whether they believe that such detrimental effects can be reversed. Data from the Survey of Health, Ageing and Retirement in Europe, which contain a numerical measure of subjective survival probability, are used to estimate a simultaneous recursive system of equations for survival expectation, subjective health and smoking. Endogeneity and unobservable heterogeneity are addressed using a finite mixture model. This approach identifies two types of individuals that differ in level of optimism, risk perception and rationality in addiction. One important result is that for both types past smokers perceive smoking consequences as reversible, with some difference between the short and long-term. We also find evidence of differences among current and past smokers in the way they evaluate the opportunity cost of tobacco consumption.

Keywords: survival expectations; subjective health; risk; smoking; EM algorithm.
JEL codes I12, C0, C30, C41.

* CRENoS, Università di Cagliari, viale S. Ignazio, 78, 09123, Cagliari, Italia. E-mail address: silvia.balia@unica.it

Acknowledgments
The author wishes to thank Andrew Jones, Casey Queen, Teresa Bago d'Uva, Martin Forster, Rinaldo Brau, Silvana Robone, Cinzia Di Novi and participants at the SIEP conference (2008) and at iHEA congress in Toronto (2011) for their suggestions and comments.

2 Introduction

In the last decade many European countries have been committed in tobacco control action plans and smoke-free legislation. Nevertheless, the failure to fully account for individual perception of the consequences of current and past smoking can likely reduce the impact of policy interventions. This motivates studying smoking behaviour and risk perception of smokers.

Measures of beliefs about survival probability can be used in models of health investments to investigate choices of unhealthy behaviours, such as smoking, over the life cycle.[1] Researchers need to be aware, however, that such measures might be largely affected by individual perception of risk, which depend on how individuals evaluate the costs and benefits associated with their behaviour. In particular, the empirical literature has provided mixed evidence on the role of smoking on perception of health and mortality risks as well as on heterogeneity in how perceived risk is related to smoking status (Gerking and Khaddaria, 2011). Viscusi (1990) support the idea that both smokers and non smokers overestimate the contribution of smoking to the onset of lung cancer. In contrast, Schoenbaum (1997) finds that heavy smokers (defined by number of cigarettes smoked) tend to underestimate the negative effect of smoking intensity on survival probability. Studying how smokers perceive their own mortality risk in response to smoking-related and general health shocks, Smith et al. (2001a, 2001b) find a significant difference between categories of smokers. Heavy smokers are more optimistic about future survival than they should be; current smokers reduce their survival expectations more dramatically than former and never smokers when they experience smoking-related health shocks. Khwaja (2007) find that, when reporting survival expectations, current smokers are very optimistic while past smokers are relatively pessimistic.

The economic literature explains smoking behaviour according to two alternative theories, which may be of help to understand risk perception of smokers. One theory defines smokers as myopic individuals who care more about present satisfaction than about the future (see Thaler and Sheffrin, 1981; Winston, 1980): tobacco consumption is a commodity that enhances instantaneous utility and its negative effects on future health and risk of dying are not taken into account. According to the theory of rational addiction, smokers are forward-

[1] Beliefs about future events are crucial to understand how individuals make economic decisions. Income and return expectations elicited from survey questionnaires, for example, have largely been used in models of consumption and saving (see, Guiso et al.,1992).

looking individuals who take into account future consequences of their decisions (Becker and Murphy, 1988), meaning that the detrimental effects of smoking are internalised. Thus, forward-looking smokers decide to smoke only if the benefits outweigh the costs of smoking, and present-oriented (myopic) individuals are potentially more addicted.

More recent works have reconciled these views. Carbone et al. (2005) define a theoretical model based on expected lifetime utility and describes how rational individuals might change their smoking behaviour depending on different perceptions of the health effects. Arcidiacono et al. (2007) investigate whether models of forward-looking behaviour explain heavy smoking (and drinking) better than models of myopic behaviour in the elderly, taking into account unobservable heterogeneity. Assuming that the myopic model can be simply nested within the forward-looking model, they show that both models predict decreasing smoking rates: smoking is less attractive as individuals age, when more illnesses occur and health worsens. Sharp declines are predicted by the fully rational model, where, however, individuals aged 50-62 years old smoke more than myopic individuals; after the age of 62 and up to the age of 80, smoking rates are higher in the myopic model; at that cut-off age of 80 there is an upward trend in smoking behaviour. This “end-of-life effect” is what Becker and Murphy would defined as a rationally myopic attitude of older people, the latter being less concerned with future effects on health. The relationship between the dynamics in smoking behaviour and age is confirmed also in other studies: Orphanides and Zervos (1995) and Kremers et al. (2004) find that the youngsters, typically subject to strong peer influences, are more likely to experiment with smoking. In addition to that, for them there is still a perception of long lifespan during which they can compensate for smoking effects by diversifying their investments in health.

This work aims to assess risk perception of smokers in reporting survival expectations and subjective health. In particular, our analysis will try to understand whether individuals believe that the effects of smoking are *reversible*. Reversibility would imply, for example, an understatement of the true effect of time spent smoking as long as smokers quit at some point in their life.[2] This is of particular concern in the case anti-tobacco campaigns pass the information that smoking is bad for your health but that quitting cancels out long-term risks. Specifically, using information on duration of the smoking habit, our analysis investigates individual perception of smoking effects in the short-term, that is on current health status (for

[2] Using numerical simulation, Carbone et al. (2005) compare two hypothetical and extreme scenarios where individual’s beliefs about the probability of death depend on health and past smoking (the *irreversible* case) and on health only (the *reversible* case).

example smoking might have effects on pulse rate, blood pression, weight), as well as in the long-term, that is on mortality risk (this is expected to be higher for smokers since the risk of onset of lung cancer or cardiovascular diseases, for example, is higher compared to the non-smoking population).

This study adds to the empirical literature on risky health behaviours in a number of ways. First, it specifies a simultaneous recursive model for individual survival expectations, subjective health and smoking behaviour that accounts for unobservable factors (such as genetics, individual rates of time preference, level of risk aversion and optimism, ability to internalise available information about health and longevity risks) that might simultaneously affect them. This is possible thanks to the estimation of a finite mixture model, via the EM algorithm, that allows identifying two homogenous individual classes in the population and class membership probabilities. Second, the empirical analysis makes use of data on elderly Europeans whilst most of the evidence on risk perception of smokers is based on US data. Third, smoking behaviour is measured as the number of years spent smoking which gives us the scope for analysis of the hazard of quitting. This enables us to shed new light on differences among types of smokers, and across classes, in reporting survival expectations and health. One important result is that in both classes past smokers perceive smoking consequences as reversible, with some difference between the short and long-term. Such attitudes in risk perception are interpretable, particularly for older smokers, in terms of myopic behaviour. Our results combines alternative ways of explaining smoking behaviour: one that tries to understand how individuals perceive the health consequences of smoking and others that are based on the traditional approach that emphasises utility maximisation of rational individuals (see Cawley and Ruhm, 2011). Finally, our analysis also shows that numerical indicators of subjective survival probability are complete measures of survival expectations that could be used in models of health investments over the life-cycle.

3 Data and variables

We use data from the first wave (2004) of the the Survey of Health, Ageing and Retirement in Europe (SHARE), a survey designed after the role models of the HRS and the English Longitudinal Study of Ageing (ELSA).[3] The target population of the survey consists of non-institutionalised individuals, and their spouses, aged 50 or older. The SHARE provides

[3]This paper uses data from SHARE release 2.4.0 supplied by the CentERdata. SHARE data collection in 2004-2007 was primarily funded by the European Commission through its 5th and 6th framework programmes. More information are available http://www.share-pro ject.org/.

rich information about health and lifestyles as well as survival expectations, which has so far been absent from European surveys. The original sample is made of 31,115 individuals, with a response rate of about 85 per cent. For the purpose of our analysis and because of item non response, this has been reduced to 20,524 respondents aged 50 to 85 years old, from northern (Denmark and Sweden), central (Austria, Belgium, Germany, France and the Netherlands) and southern Europe (Italy, Spain and Greece).

3.1 Indicator of survival expectations

Survival expectations are measured by a numerical indicator of subjective survival probability (SSP). This is derived from the question "What are the chances that you will live to be age T or more?". The distribution of SSP is approximated by proposing a different target age (T) to each individual depending on her age class. We consider target ages of 75, 80, 85, 90 and 95, where the distance from current age is between 14 and 25 years (see Table 1). We exclude individuals older than 85 years old, because of low frequency in the sample, as well as those who were not matched to the appropriate target age.

TABLE 1

The question on SSP is the ninth of eleven questions about predicted probabilities of future events. Responses are driven by a card reporting a sequence of numbers from 0 ("absolutely no chance") to 100 ("absolutely certain"). A warm-up question, "What do you think the chances are that it will be sunny tomorrow?", is asked to help respondents feel at ease with probabilities. Nevertheless, we excluded records with unreliable responses according to a criterion based on two questions about the chance that the standard living will be better and worse. A subjective probability of 0 for both events indicates a high expected probability that the standard living will be unchanged. If the probabilities sum to greater than 1, this would indicate that subjective assessment of future events is unreliable.[4]

Eliciting survival expectations through predicted probabilities is usually preferred to asking qualitative questions: probabilities permits a better comparison across individuals, while qualitative responses may depend on cognitive, linguistic and cultural differences and usually suffer from response bias. Furthermore, internal consistency of probabilities can be assessed (Juster, 1966; Dominitz and Manski, 1997; Manski, 2004). The main disadvantage of using using SSP, however, is heaping at focal values such as 0, 50, 100 an values ending with a zero. Figure 1 shows the distribution of SSP at each target age and for each European region

[4] We use a tolerance level of 0.10 and exclude only individuals for whom the sum of the probabilities is higher than 1.10.

considered. Around 4.1 per cent of all respondents report to have no chance to survive to the target age, 25.2 per cent report a SSP of 50 and 15.8 per cent are certain to survive. Responses are more concentrated at higher values, as 60, 70, 70 and 90. We deal with this issue in the econometric modelling by choosing an appropriate parametric distribution for SSP.

Another advantage of using a quantitative measure of expectations is that the distribution of SSP is comparable with either probabilities computed from life tables or observed mortality. Life tables are usually found to understate survival probability and SSP is broadly considered a better predictor of future mortality than objective life table hazard rates.[5] A comparison of SSP from the SHARE to actuarial survival probabilities from life tables, by Peracchi and Perotti (2010), confirm this view and show that life tables omit important individual characteristics while subjective probabilities may be affected by the individual's level of optimism. Figure 1 also shows compares average SSP with average probabilities of survival calculated from period life tables for 2004, available in the Human Mortality Database. As expected, average SSP decreases with target age. Southern Europeans report, on average, higher probabilities than the others, except for survival at 75 where Northern Europeans are more optimistic (SSP is 72 per cent). SSP at 75, 80 and 85 are always lower that those calculated from life tables; the opposite is true for survival at 90 and 95. This might capture the fact that SSP likely depends upon observable (socioeconomic status, health and risk factors) as well as unobservable individual characteristics influencing beliefs (such as level of optimism), which are not included in life tables.

FIGURE 1

3.2 Measures of subjective health and smoking behaviour

A commonly used measure of general health status is self-assessed health (SAH) (see Deaton and Paxson, 1998; Smith, 1999): it measures perceived health but is known to be a good indicator of morbidity and a powerful predictor of future health and mortality (Idler and Benyamini, 1997; van Doorslaer and Gerdtham, 2003). Respondents are asked "How is your health?" and can answer "excellent", "very good", "good", "fair", "poor". We use a

[5] Hurd and McGarry (1995) show that SSP is internally consistent if compared to the average survival probability from life tables and covaries well with socio-economic variables and risk factors as actual mortality does. Hurd et al. (1999) and Hurd and McGarry (2002) stress the role of unobservable heterogeneity in SSP, whereas life tables only capture the effect of demographic factors on mortality. Despite being correlated with health and the onset of diseases, SSP is not simply an alternative measure of overall health status: it has an element of expectations that accounts for most of the unobservable heterogeneity, understanding which would help in the estimation of life-cycle models.

dichotomised version of SAH that takes the value 1 if health is excellent or very good, and 0 otherwise.[6]

The SHARE also provides rich information about tobacco smoking habit. On the basis of the questions, “Have you ever smoked cigarettes, cigars, cigarillos or a pipe daily for a period of at least one year?”, “Do you smoke at the present time?”, it is possible to know whether respondents never smoked, are current smokers or have quit by the time of the interview (past smokers). Combining this information to responses at the question “How old were you when you stopped smoking?” we build a duration time variable, indicating the number of years spent smoking. This variable is right-censored at the time of the interview for current smokers who can eventually quit; complete spells of smoking are observed only for past smokers.

3.3 Descriptive statistics

Table 2 reports sample means for the variable used to describe the sample. Average SSP is 61.9 per cent; 32 per cent of respondents are in excellent or very good health. Current smokers counts for 20 per cent of the sample and have already smoked for 36.5 years on average. About 28.6 per cent of smokers have quit by the time of the interview and have smoked for 22.5 years.

Table 2 also compares individuals whose SSP is equal to 50 per cent with those who report lower and higher probabilities. A 50 per cent SSP might reflect “epistemic uncertainty” rather than probabilistic thinking (Bruine de Bruin et al., 2002). This should not preclude, however, the possibility that such response is a genuine survival expectation (Hill et al., 2005). The proportion of respondents in excellent or very good health increases monotonically moving from the group with SSP lower than 50 to the group with the highest SSP. The same trend is found in the number of disabilities. The proportion of sedentary and obese individuals decreases moving from the group with lower to that with higher SSP; while the proportion of drinkers and smokers increases. Current smokers, however, are more concentrated in the 50 and higher SSP groups, while past smokers in the highest probability group. As expected, longer smoking duration are associated with lower probability. Individuals in poor socio-economic status are concentrated in the group with lower SSP. For those who report a SSP lower than 50 more mothers, fathers and spouses have already died with respect to the other two groups, where the age at death of parents is , in fact, slightly higher. On the contrary, higher SSP is associated with lower age at death of the spouse. Overall, observable

[6] Etilè and Milcent (2006) find that a way to reduce reporting heterogeneity in SAH is to convert the ordered variable in a binary variable.

differences across subgroups suggests that a 50 per cent SSP is likely to be a genuine response.

TABLE 2

3.4 Variation with health, smoking, socioeconomic variables and parental mortality

Observed variation in SSP should be in line with epidemiological evidence on the relation between mortality risk, health status, smoking and socioeconomic status. Table 3 shows that, survival expectations get dramatically lower moving from excellent to poor health and decrease as individuals age (that is, for higher target ages). Surprisingly, average SSP at ages 75, 80 and 85 is higher for past relative to current and never smokers, while the lowest probabilities is reported by current smokers; average SSP at age 90 is higher for current and never smokers. If we look at the overall distribution, past and current smokers report the highest average SSP. As expected, SSP is positively related to income and education.

TABLE 3

Table 4 shows that the first and second quartiles of the SSP distribution are mainly composed by individuals in good and fair health; the second quartile also includes a larger proportion of individuals in very good health. The last two quartiles show larger variation in the distribution of SAH and include fewer individuals in fair health. Never smokers, however, represent about half of individuals in each SSP quartile and this proportion decreases monotonically moving from the first to the last quartile. The composition of each SAH category by smoking status is very alike. Never smokers are more concentrated in the fair category, past smokers in the excellent, very good and poor categories, and current smokers are evenly spread in the five categories. Smoking durations is longer in the first and second SSP quartile, and increases moving from the excellent to the poor health group.

TABLE 4

Other studies found a relation between parents' death and SSP (Hurd and McGarry, 1995, 2002). With some exception, average SSP is always 10 per cent lower if one of the parents is dead by the time of the survey (Table 5). The relation between age at death of parents and SSP is less clear. One explanation is that parents' early deaths are not related to children's survival but depend, in turn, on accidents. SSP is also higher if the spouse is still alive or died within her fifties, declines for death between ages 51 and 74, and reaches the lowest level when death occurs at the oldest ages.

TABLE 5

4 Methods

Our empirical model is based on the idea of health investments, as in Grossman (1972), and further assumes that expected lifetime utility of a rational individual depends on the instantaneous utility derived by her own health, consumption of tobacco and other commodities; utility is discounted by intertemporal preferences and beliefs about survival, as in Carbone et al. (2005).

This takes the form of a simultaneous recursive model for survival expectations (sE_i), subjective health status (sH_i) and smoking behaviour (S_i); where sE_i at any specific age depends on sH_i. The equation for sE_i considers that, assessing survival, individuals weigh up the direct effects of smoking on mortality risk as well as the indirect effects through health (long-term effects of smoking) and is written as:

$$sE_i = f(sH_i, S_i, X_i, \mu)$$

Subjective health depends on objective health (H_i), is influenced by perception of the direct effects of smoking on current health (short-term effects of smoking), and is expressed by the following health production function:

$$sH_i = f(H_i, S_i, X_i, \mu)$$

Smoking behaviour is defined as:

$$S_i = f(X_i, \mu)$$

The three processes also depend on individual observable characteristics (X_i) and unobserved factors μ. Health and smoking can be endogenous to survival expectations and, in turn, smoking can be endogenous to health. This might be due to unobserved factors which simultaneously affect formation of expectations, reporting health and smoking behaviour such as genetics, individual rates of time preference, level of risk aversion and optimism, ability to internalise available information about health and longevity risks.

4.1 Model specification

We approximate the distribution of SSP using the beta distribution, often employed for proportions or subjective beliefs about probabilities of future events (Ferrari and Cribari-Neto, 2004; Smithson and Verkuilen, 2006), because it models well continuous and bounded variables characterised by spikes at certain response foci:

$$f(sE_i|x_1) = \frac{\Gamma(\omega+\tau)}{\Gamma(\omega)\Gamma(\tau)} {y_1}^{(\omega-1)}(1-y_1)^{(\tau-1)} \quad (1)$$

where sE_i is rescaled SSP, Γ is the gamma distribution, and ω and τ are shape parameters.[7] Maximum likelihood (ML) estimation techniques can be used.[8] The expected value of *SSP* is approximated by a logistic: $E(SSP) = \frac{\omega}{\omega+\tau} = \frac{\exp(x_1\beta)}{1+\exp(x_1\beta)}$ where, for simplicity of notation, x_1 includes subjective health, smoking, socioeconomic and demographic variables.[9]

Other controls used to explain survival expectations are parental and spouse mortality, which have been largely employed in previous works (see, for example, Hurd and McGarry, 2002). To control for systematic differences in reporting expectations and since respondents are not asked to evaluate chances of survival for the same number of years (as shown in Table 1) x_1 also includes age classes corresponding to target ages as well as a continuous indicator of the difference between individual age and target age. Given the very nature of SSP, an indicator of numeracy that captures cognitive ability is included as well.

For subjective health we use a probit model which describes the probability of being in good or very good health:

$$\Pr(sH_i = 1|x_2) = \Phi(kx_2\beta) \qquad (2)$$

where sH_i is SAH, $k = (2y_2 - 1)$ is an indicator of sign and x_2 are individual characteristics including smoking, socioeconomic and demographic variables. Other controls used to better identify the causal relation between SAH and SSP are indicators of objective health such as disability (limitations in usual activities because of health problems, gali; in activities of daily living, adl; in instrumental activities of daily living, iadl) and hospital admission in the last twelve months. It is reasonable to think that these indicators have a direct effect on reporting health but do not affect directly beliefs about survival, since individuals usually adapt quickly to long-term disability and sudden health changes.

Smoking is modelled using a two-part specification of the duration model, which implies splitting the sample according to starting (Douglas and Hariharan, 1994):[10]

$$[\Phi(x_3\beta)f(t|x_3,\beta)]^{s\cdot q}\,[\Phi(x_3\beta)S(t|x_3,\beta)]^{s\cdot(1-q)}[1-\Phi(x_3\beta)]^{(1-s)} \qquad (3)$$

---

[7] We rescale y_1 to lie in the (0, 1) interval. To avoid the logarithm of zeros and ones, $y = 100\ (N-1) + a$, where N is the sample size and a is some constant (here $a = 0.5$). Alternative transformations can be used.

[8] One common approach is to convert the dependent variable using the logistic transformation and then use the normal distribution for estimation. Paolino (2001) shows that ML estimation using the beta distribution may provide more accurate and more precise results than OLS. This requires reparametrising the likelihood function such that a location and a precision sub-model are defined.

[9] Smoking indicators are two dummy variable for current and past smokers and their interactions with smoking duration. The logarithm transformation is used to ensure flexibility.

[10] Other applications can be found in Douglas (1998), Forster and Jones (2001) and Balia and Jones (2011).

where t is the duration of smoking, s is a binary indicator that takes value 1 if the individual eventually started smoking, q is a binary indicator that takes value 1 if the individual quit.[11] A probit for the probability to start, $\Pr(\mathrm{s}=1|x_3)=\Phi(x_3\beta)$ describes the first part of the model; x_3 includes only income, education and demographic variables, as they are assumed to reflect past socioeconomic characteristics influencing starting. The second part models the hazard of quitting; it follows a Weibull distribution, with density $f(t|x_4)=\lambda\alpha t_i{}^{(\alpha-1)}\exp\left(-\lambda t_i{}^{\alpha}\right)$ and survival $S(t|x_4)=\exp\left(-\lambda t_i{}^{\alpha}\right)$.[12] Here α is the duration dependence parameter and λ is a function of covariates $\exp\left(-x_4\beta\right)$, where x_4 includes socioeconomic and demographic variables.[13] To better identify the causal effect of smoking on subjective health and survival expectations indicators of drinking, physical exercise and obesity are included. Sedentary behaviour, alcohol consumption and dietary habits are strictly correlated with smoking behaviour (Marcus et al., 1999; Di Novi, 2010). An indicator of smoking-related diseases diagnosed before the interview and an indicator of household composition (the number of children) are also included. These variables might influence directly the hazard of quitting, with only indirect effect on SAH and SSP.

Taking together equations (1) to (3), the sample likelihood of our recursive model is:

$$L_i=f(y_1|y_2,t,s,q,x_1)\cdot\Pr(y_2=1|t,s,q,x_2)\cdot[f(t|x_4)\Pr(\mathrm{s}=1|x_3)]^{s\cdot q}$$
$$\cdot\,[S(t|x_4)\Pr(\mathrm{s}=1|x_3)]^{s\cdot(1-q)}\cdot[1-\Pr(\mathrm{s}=1|x_3)]^{(1-s)} \qquad (4)$$

In the presence of unobservable heterogeneity (μ), omitted for simplicity of notation, this is analytically intractable and an appropriate estimation approach is needed.

4.2 Estimation approach

We propose a finite mixture model, which divides the population in a finite number of individual types or latent classes from which the observed data come from (McLachlan and Peel, 2000). Response variables are assumed to be independent of one another given class membership of each individual. If this holds, a single response per individual is sufficient to identify the model with cross-sectional data. Finite mixture models have been recently used in

[11] This distinguishes between never ($s=0$; q is not observed), past ($s=1$ and $q=1$) and current smokers ($s=1$ and $q=0$).
[12] The Weibull is the most appropriate parametrisation for the hazard of quitting smoking (see Douglas, 1998). In this application, it has been chosen among other distributions on the basis of information criteria (AIC and BIC) and Cox- Snell residuals test.
[13] λ is non-negative because the model is parameterised in the accelerated failure time metric. Estimated coefficients are to be interpreted in the terms of acceleration (or deceleration) of time to quitting.

applications to cross-sectional data with count (see e.g., Deb and Trivedi, 1997, 2002, 2006) and binary (see e.g., Atella et al., 2004) responses.

ML estimation of the mixture for equation (4) is achieved using the EM algorithm, usually appropriate in the presence of incomplete or missing data (Dempster et al., 1977). Using general notation, equation (4) can be written as $f(y_i|x_i;\Theta) = \sum_{k=1}^{K} p_k \cdot f_k(y_i|x_i;\theta_k)$ where $\Theta = (p_1, \dots, p_K, \theta_1, \dots, \theta_K)$ and f_k is a density function parameterised by θ_k. The expression above assumes that there are K component densities mixed together; each p_k is a mixing proportion and can be interpreted as the prior probability ($0 < p_k < 1$ and $\sum_{k=1}^{K} p_k = 1$) that a data point is randomly drawn from component k. The unconditional sample log-likelihood for the incomplete data is difficult to maximize because it contains the logarithm of a sum, and unknown mixing parameters:

$$logL(y_i|x_i;\Theta) = log\prod_{i=1}^{N} f(y_i|x_i;\Theta) = \sum_{i=1}^{N} \log\left(\sum_{k=1}^{K} p_k \cdot f_k(y_i|x_i;\theta_k)\right) \qquad (5)$$

Using Bayes' rule, we specify the posterior probability $\hat{\pi}_{ik}$ of class membership, conditioned on covariates and the outcomes:

$$E(\pi_{ik}|y_i, x_i;\Theta) = \frac{p_k \cdot f_k(y_i|x_i;\theta_k)}{\sum_{k=1}^{K} p_k \cdot f_k(y_i|x_i;\theta_k)} = \frac{p_k \cdot f_k(y_i|x_i;\theta_k)}{f(y_i|x_i;\Theta)} = \hat{\pi}_{ik} \qquad (6)$$

The estimated unconditional probability p_k from equation (5) is the mean of the conditional probability $\hat{\pi}_{ik}$: $\hat{p}_k = E(\hat{\pi}_{ik}) = \frac{1}{N}\sum_{i=1}^{N} \hat{\pi}_{ik}$. After manipulation and substituting p_k in equation (5) with its value derived from equation (6), the conditional log-likelihood is:

$$logL\left(y_{i,}\Pi_i\middle|x_i;\Theta\right) = \left[\sum_{i=1}^{N}\sum_{k=1}^{K} \hat{\pi}_{ik} \cdot log\, f_k(y_i|x_i;\theta_k) + \sum_{k=1}^{K} \hat{\pi}_{ik} \cdot log(p_k)\right]$$
(7)

Estimates for θ_k are obtained by $argmax \sum_{i=1}^{N}\sum_{k=1}^{K} \hat{\pi}_{ik} \cdot \log f_k(y_i|x_i;\theta_k)$ and so the values of $\hat{\theta}_k$ maximise both the unconditional and the conditional likelihood. Intractability of the log-likelihood is overcome because the function to be maximised has again an additive form.

The EM iterates two steps until convergence. The E-step computes the conditional expectation of the expression $\log f_k(y_i|x_i;\theta_k)$ and $\hat{\pi}_{ik}$ according to equation (6). The M-step is simply implemented as the ML estimation of weighted models, using the posterior probabilities as weights. The two steps alternate in a loop that starts with initial values for the parameters (p_k^*, θ_k^*). At the first iteration (r) the E-step calculates $\hat{\pi}_{ik}^{r} = Pr\,(\Pi_{ik} = 1|y_i, x_i;\theta^*, p^*)$. The M-step provides the updating formulas $p_k^{r+1} = \frac{1}{N}\sum_{i=1}^{N} \pi_{ik}^{r}$ and $\theta_k^{r+1} = argmax \sum_{i=1}^{N}\sum_{k=1}^{K} \hat{\pi}_{ik} \cdot \log f_k(y_i|x_i;\theta_k)$ that are used to compute $\hat{\pi}_{ik}^{r+1}$. The likelihood increases monotonically at each iteration. Under suitable regularity conditions, the sequence

θ^r converges to a stationary point of $L(\theta)$.[14] Properties of convergence, including these conditions, are discussed in Dempster et al. (1977), McLachlan and Krishnan (1996) and Schafer (1997).

The number of latent classes (K) is usually chosen according to information criteria. Typically two or three classes are chosen. We present results from a model with two classes: this seems to provide a natural representation of our data given that it helps distinguishing between subgroups of more and less optimistic individuals.

5 Results

Table 6 shows that, according to information criteria, such as the AIC, BIC and CAIC the mixture model with two latent classes performs better than the single class model, thus providing evidence of unobservable heterogeneity. The estimated parameters p_1 = 0.66 and p_2 = 0.34 are the probabilities that an observation is randomly drawn from the first and the second latent class. Slope parameters are allowed to vary across classes and for each equation, as reported in Table 7.

TABLE 6

5.1 Survival expectations

Estimated beta regression coefficients can be transformed in odds ratios (see Table 8): this makes it possible to interpret percentage changes from the average SSP of the baseline individual.[15] This is defined as a single female aged between 81 and 85, at average distance from target age, comes from a southern European country, is in good or less than good health, never smoked, has low income and no education, is unemployed (or a housekeeper), her parents are still alive, and has average cognitive ability. In class 1 the baseline individual expects to have a 75.4 per cent probability of survival at age 95. The same baseline individual has lower probability, 47.5 per cent, in the second latent class, thus suggesting that the two types of individuals are characterised by a different level of optimism. As expected, average SSP decreases in both classes when the distance between age and target age increases and is

[14] The likelihood of a mixture model might not be unimodal, meaning that there are several local maxima and a unique global maximum. The solution found by the EM loop can depend critically upon the set of initial values for the prior probabilities and the θs. One possibility is to run the loop several times with different initialisations and choose the best model comparing the likelihood values. Initial values can be guessed, or can be computed as a linear transformation of the parameter estimates from the single component model.
[15] This calculation only requires exponentiating the coefficients. The odds ratio is preferred to average partial effects because the latter measure a percentage point change, while here we want to measure a percentage change.

higher in each younger age class: in particular, in class 1 average SSP is 21.5 per cent higher for those who evaluate survival at age 90; 148 and 112 per cent higher for those who evaluate survival at age 75 and 85 respectively. This is amplified for individuals in the second latent class, where average SSP is 331 and 221 per cent higher for those who evaluate survival at age 75 and 85.

TABLE 7 – 8

The role of socioeconomic variables varies across latent classes. In class 1, individuals who belong to the second poorest income quartile expect lower SSP (-3 per cent), while those in the highest income quartile expect higher SSP (+6 per cent) relative to the baseline. In class 2, the effect of income seems more important: being in the third (near rich) quartile implies an increase of 13 per cent in baseline SSP. Education matters only in class 1: qualifications lower than diploma decrease average SSP of 8 per cent. Employed individuals expect to live longer: average SSP is 26.5 per cent higher in class 2 and 10 per cent in the first one. In class 2 also retirement positively affect expectations (+9 per cent). In class 1, average SSP increases of about 7 per cent for either married or separated individuals; in class 2, the contribution of marriage is substantial and close to that of SAH (+41 per cent). Also widows expect higher SSP (+ 27 per cent) but there is no effect of age at death of the spouse. Average SSP is significantly associated to father's mortality: it decreases of about 13 per cent if father died. In class 1, however, it increases significantly for ages at death older than 50. Mather's mortality matters only in class 1 where average SSP is 10 per cent lower if the mother died, but increases of about 4 per cent for ages at death higher than 80. The indicator of numeracy is statistically significant only in class 2, where average SSP is 5 per cent higher as cognitive ability increases. Regional dummies have a significant and larger effect in class 2, where average SSP decreases of 23 and 28 per cent if individuals come from northern and central Europe respectively (12 and 10 per cent in class 1).

As expected, we find that SAH explains most of the observed variation in SSP. Its effect is predominant in class 2, where being in excellent or very good health increases average SSP of about 54 per cent (about 37.6 per cent in class 1).

In class 1, average SSP significantly increases for past smokers (+16.6 per cent) and quitting compensates about 4 years of smoking. This counterintuitive result can be interpreted in terms of myopic behaviour: those who quit do not internalised the negative effect of past smoking on mortality risk, rather they reward themselves for quitting with better chances of living longer than their smoking behaviour would warrant. This suggest that long-term effects are perceived as reversible. Surprisingly, indicators of current smoking habit are not significant, although the signs of the coefficients show that, relative to never smokers, current smokers expect to live shorter. In class 2, none of the smoking indicators is statistically significant and both past and current smokers expect a slightly higher survival for an additional year of smoking (1.3 and 0.8 per cent respectively). Overall, our results suggests that when smoking matters (in class 1), there is still a substantial difference between current and past smokers. Current smoking does not significantly affect formation of survival expectation, past smoking duration has some negative effect on future mortality risk but this is balanced out by quitting.

5.2 Subjective health

We estimate average partial effects (APE) of covariates on the probability of reporting good or excellent health.[16] Table 8 shows that the likelihood of reporting better health is higher for individuals in the richest income quartile and for those who are employed; APEs of these variables are bigger in the second latent class. In both classes the more educated report better health, but having a diploma and a university or higher degree has a bigger effect in the first latent class. In class 2 marital status is important: married, widows and separated or divorced report better health than singletons. The probability of reporting good health decreases with age: in class 1 it is significantly higher only for the youngest old (+5.8 per cent), while in class 2 it is about 10 per cent higher for individuals aged 50-65 and 6 per cent for those aged 66-70. Disability indicators have the expected effect: the probability of reporting better health decreases with the number of limitations. Adl and iadl have a larger impact in class 2 (+14 and 7 per cent); gali in class 1 (-30 per cent). Those who have been hospitalised in the last year tend to report worse health: this effect is larger in class 2. Region of origin matters more for class 2, where the probability of reporting better health is 21 and 3 per cent higher for the northern and central Europeans.

Smoking does not have a significant effect on SAH in class 1, while in class 2 the probability of reporting better health is significantly higher for past smokers (7.7 per cent higher than never smokers) and decreases of about 2.9 per cent for one additional year spent smoking. This can be interpreted again in terms of myopic behaviour and reversibility. Indicators of current smoking are not significant and the APEs are very close to that of past smoking. These results suggest that smoking affects subjective health only in the

---

[16] Partial effects are computed for each individual as the change in the probability that SAH equals 1 when a covariate changes, then averaged across the whole sample, so that they refer to the entire population. We use the finite difference method for dummy variables and the calculus method for continuous variables as in Wooldridge (2002).

subpopulation of the second type (the same for which smoking does not affect expected survival). They also confirm a substantial difference between types of smokers.

5.3 Smoking behaviour

As shown in Table 7, the propensity to become a smoker is higher for men, more educated individuals as well as for those from the youngest age cohorts in both classes. Income is positively related to starting but the gradient of the relationship is not clear. Our model reflects heterogeneity in the hazard of quitting. Smoking duration is predicted to be significantly shorter for richer, retired and employed smokers, particularly in class 1. Educated individuals smoke longer relative to those without education, particularly in class 2. In both classes smokers with a unhealthy lifestyle, in terms of drinking and exercise, tend to quit later, while those who have a bad dietary habit quit earlier. Northern and central Europeans are predicted to quit earlier. The two classes differ, instead, in the way tobacco-related diseases affect quitting: only in class 1 smokers who report diseases tend to quit earlier. Married individuals are predicted to quit earlier than others, especially in the second latent class. The predicted probability of starting is lower in class 1 (0.47; 0.51 in class 2), while the estimated average smoking duration does not differ much between classes (40.8 and 40.6 years).

5.4 Predicted survival expectations and subjective health for types of smokers

Table 9 shows that predicted average SSP is higher in class 2 (65.2 per cent, 58.1 per cent in class 1). However, predicted average probability of SAH=1 is very alike across classes (32.2 and 32.6 per cent). We would have expected to find, instead, better health in the class with higher expectations.

TABLE 9

The table also compares, for each latent class, predicted survival expectations and subjective health calculated for the overall sample and hypothetical scenarios, defined on the basis of smoking behaviour and target age. This makes it possible to discuss differences in risk perception between never, former and current smokers. The first two columns of Table 10 show that predicted average SSP is always higher in class 2; the highest expectation (61 and 66 per cent) is associated to the scenario where all individuals quit smoking. In class 1, survival expectations is about 3.6 and 4.5 percentage points lower in the “smoking free scenario” (individuals never smoked) and in the “smoking scenario” (all individuals currently smoke). The difference between the “quitting scenario” and “smoking free scenario” is negligible (0.6) in class 2, while survival expectations is about 3.6 percentage points lower in the “smoking scenario”. Comparing the “smoking free scenario” with the “smoking scenario” we find that expectation is higher in the first case, with a 0.9 and 2.6 percentage points difference in class 1 and 2 respectively.

Predicted probability of SAH=1 is always lower in the “smoking free scenario”. The highest probabilities are associated to the “quitting scenario” and the “smoking scenario”, with a difference of about 4 and 7 percentage points with the “smoking free scenario” in class 1 and 2 respectively.

Results confirm that risk perceptions of smokers is altered by the decision to quit for what concerns both the short and the long-term effects of smoking. Risk perception of current smokers seems to take into account the long-term but not the short-terms effect of smoking.

5.5 Posterior analysis

Estimation shows that our two individuals types behave differently in the way they formulate survival expectations and report subjective health. For a deeper investigation we examine the determinants of class membership. We assign each individual to the class associated to the larger posterior probability using a cut-off probability of 0.5 (see Atella et al., 2004), and define a binary indicator of class membership that takes value 1 if the posterior probability is above the cut-off probability and 0 otherwise. This is equivalent to saying that individual *i* belongs to latent class 1 if the estimated posterior probability π_{i1} is larger than the estimated π_{i2}, since $\sum_{k=1}^{K} \pi_{ik} = 1$.

Table 10 shows that, on average, sample 1 (drawn from the first sub-population) reports a lower probability of survival at some future age, about 59 per cent (about 70 per cent in sample 2). In both samples about 32 per cent of respondents report very good or excellent health, and this proportion is slightly lower in class 2 where, however, a higher percentage of individuals reports gali limitations and a higher number of disabilities (adl and iadl). Smokers are more concentrated in sample 2 (47.8 per cent of respondents never smoked in sample 2; 52.4 in sample 1). Of them, 21.7 per cent smoked at the time of the interview while 30.5 had already quit (19.6 and 28 per cent in class 1); average time spent smoking is about one year longer in the second sample. Sample 2 appears to represent a population of “more optimistic” individuals who are also “less healthy” and “more addicted to smoking”. It is also characterised by relatively older individuals (it includes a larger share of individuals aged 71 and over than sample 1), most of them are retired or belong to the poorest income quartiles, have lower education and a less healthy lifestyle.

TABLE 10

We additionaly estimate a probit model for class membership conditional on the exogenous variables used in the mixture model (Table 11) The probability of being assigned to class 1 (where "less optimistic", "healthier" and "less addicted to smoking" individuals concentrate) is significantly lower for retired, married or separated individuals, those who have been in hospital in the last year, and have an unhealthy lifestyle. It decreases with the number of children and increases with cognitive ability. Class membership significantly depends on age: individuals aged 50-79 are more likely to appear in class 1, meaning that those approaching the end of lifespan might find smoking still attractive. According to the literature, we interpret this in terms of rational myopic behaviour: smoking is a rational choice when individuals believe that there is not enough time left for smoking to affect mortality risk, meaning that the associated opportunity cost is low. This finding is in line with Adda and Lechene (2001): they present a joint model of tobacco consumption and mortality over the life-cycle where individuals with lower life expectancy select into smoking, the cost of smoking being higher for those with longer potential life expectancy.

TABLE 11

6 Discussion

This paper explores formation of survival expectations focusing on the role of smoking on people's perception of health risk. The empirical literature, mainly based on US surveys, provides mixed evidence on the relation between smoking and risk perception.

We propose a simultaneous recursive model for survival expectation, subjective health and smoking duration. The analysis controls for endogeneity of the health and smoking variables as well as unobservables, by means of a finite mixture model. Using data on elderly Europeans, we provide evidence of heterogeneity in assessing survival probability and reporting health and identify two individual types in the population, which differ in observed characteristics as well as in the level of optimism, risk perception and rationality in addiction. The first type does not evaluate neither the short (i.e., current health damages due to smoking) nor the long-term effects of smoking (i.e., higher mortality risk due to smoking): in particular, past smokers reward themselves for quitting by reporting better chances of living longer than their smoking behaviour would warrant. For individuals of the second type, long-term effects of smoking are neglected while short-terms effects are taken into account; current smokers do not seem to discount future consequences on mortality risk. Furthermore, in reporting health status, the effect of smoking duration is once again compensated by the decision to quit.

Particularly for older individuals, such attitudes in risk perception might depend on a lower opportunity cost of smoking due to lower life expectancy.

Policy makers who are concerned with the prevention of health problems and the promotion of healthy lifestyles might be interested in knowing whether and to what extent individuals understand morbidity and mortality consequences of smoking. This paper shows that, despite existing national anti-smoking campaigns, smokers are not necessarily fully aware of the true health risk relative to the non-smoking population. This raises a first question of whether more dissemination of information is needed. Reversibility of the smoking effects reflects myopic behaviour of past smokers. This raises a second question of whether better dissemination of information would be necessary.

Our analysis also shows that the numerical indicator of subjective survival probability is a complete measure of survival expectations which captures observable as well as unobservable factors influencing individual beliefs. The use of such indicator in models of health investments over the life-cycle might be preferred to mortality hazard rates estimated from life tables, thus giving ample scope for future research.

7 References

Lechene, V. and J. Adda. 2001. Smoking and Endogenous Mortality: Does Heterogeneity in Life Expectancy Explain Differences in Smoking Behavior? Economics Series Working Papers 077, University of Oxford, Department of Economics.

Arcidiacono, P., H. Sieg, and F. Sloan. 2007. Living rationally under the volcano? An empirical analysis of heavy drinking and smoking. *International Economy Review* 48(1), 37-65.

Atella, V., F. Brindisi, P. Deb, and F. Rosati. 2004. Determinants of access to physicians services in Italy: a latent class seemingly unrelated probit. *Health Economics* 13(7), 657–668.

Balia, S. and A. Jones. 2011. Catching the habit: a study of inequality of opportunity in smoking-related mortality, *Journal of The Royal Statistical Society: Series A* 174(1), 175–194.

Becker, G. S. and K. Murphy. 1988. An theory of rational addiction. Journal of Political Economy 96 (4), 675–700.

Börsch-Supan, A., A. Brugiavini, H. Jürges, J. Mackenbach, J. Siegrist, and G. Weber. 2005. Health, Ageing and Retirement in Europe First Results from the Survey of Health, Ageing and Retirement in Europe. Mannheim: MEA.

Bruine de Bruin, W., P. S. Fischbeck, N. A. Stiber, and B. Fischhoff. 2002. What number is "fifty-fifty"?: Redistributing excessive 50% responses in elicited probabilities. *Risk Analysis* 922 (4), 713–723.

Carbone J.C., S. Kverndokkb, O. J. Røgebergb. 2005. Smoking, health, risk, and perception. *Journal of Health Economics* 24, 631-353.

Cawley J. and C. J. Ruhm. 2011. The Economics of Risky Health Behaviors. NBER Working Paper Series, w17081.

Chapman, S., W. Wong, and W. Smith. 1983. Self-exempting beliefs about smoking and health: differences between smokers and ex-smokers. *American Journal of Public Health* 83 (2), 215–219.

Deaton, A. S. and C. Paxson. 1998. Ageing and inequality in income and health. *American Economic Review, Papers and Proceedings* 88, 248–253.

Deb, P. and , P. K. Trivedi. 1997. Demand for Medical Care by the Elderly: A Finite Mixture Approach. *Journal of Applied Econometrics* 123, 313-336.

——————— 2002. The structure of demand for health care: latent class versus two-part models. *Journal of Health Economics* 214, 601-625.

——————— 2006. Specification and simulated likelihood estimation of a non-normal treatment-outcome model with selection: application to health care utilization. *Econometrics Journal* 9, 307–331.

Dempster, A. P., N. Laird, and D. B. Rubin. 1977. Maximum likelihood from incomplete data via EM algorithm. *Journal of the Royal Statistical Society. Series B* 34 (1), 1–38.

Di Novi C. 2010. The influence of traffic-related pollution on individuals' life-style: results from the BRFSS. *Health Economics* 19, 1318-1344.

Dominitz, J. and C. F. Manski. 1997. Using expectation data to study subjective income expectations. *Journal of the American Statistical Association* 92, 855–867.

Douglas, S. and G. Hariharan. 1994. The hazard of starting smoking: Estimates from a split population duration model. *Journal of Health Economics* 13 (2), 213–230.

Etilè, F. and C. Milcent. 2006. Income-related reporting heterogeneity in self-assessed health: evidence from France. *Health Economics* 15, 956–981.

Ferrari, S. and F. Cribari-Neto. 2004. Beta regression for modelling rates and proportions. *Journal of Applied Statistics* 31 (7), 799–815.

Forster, M. and A. M. Jones. 2001. The role of tobacco taxes in starting and quitting smoking: duration analysis of British data. *Journal of the Royal Statistical Society: Series A* 164, 517–547.

Gerking, S. and R. Khaddaria. 2011. Perceptions of health risk and smoking decisions of young people. *Health Economics*. DOI: 10.1002/hec.1760

Grossman, M. 1972. On the Concept of Health Capital and the Demand for Health. *Journal of Political Economy* 80 (2), 223-255.

Guiso, L., T. Jappelli, and D. Terlizzese. 1992. Earnings uncertainty and precautionary saving. *Journal of Monetary Economics* 30, 307–337.

Hill, D., M. Perry, and R. Willis. 2005. Estimating knightian uncertainty from survival probabilities on the HRS. Manuscript, University of Michigan.

Hurd, M., D. McFadden, and A. Merrill. 1999. Predictors of mortality among the elderly. NBER Working Paper 7440.

Hurd, M. and K. McGarry. 1995. Evaluation of the subjective probabilities of survival in the Health and Retirement study: Data quality and early results. *The Journal of Human Resources* 30, 268–S292.

——————— 2002. The predictive validity of subjective probabilities of survival. *The Economic Journal* 112, 966–985.

Idler, E. L. and Y. Benyamini. 1997. Self-rated health and mortality: community studies. *Journal of Health and Social Behavior* 38 (1), 21–27.

Khwaja, A., F. Sloan and S. Chung. 2007, The relationship between expectations and behaviours: mortality expectations and smoking decisions. *Journal of Risk and Uncertainty* 35, 179-201.

Juster, T. 1966. Consumer buying intentions and purchase probability: an experiment in survey design. *Journal of the American Statistical Association* 61, 658–696.

Kremers, S. P., A. Mudde, and H. D. Vries 2004. Model of unplanned smoking initiation of children and adolescents: an integrated stage model of smoking behavior. *Preventive Medicine* 38, 642–650.

Marcus B.H., B.A. Lewis, J. Hogan, T.K. King, A.E. Albrecht, B.Bock, A.F. Parisi, R. Niaura, D.B. Abrams. 1999. The efficacy of exercise as an aid for smoking cessation in women: a randomized controlled trial. *Archives of Internal Medicine* 159: 1229–1234.

McLachlan, G. J. and T. Krishnan. 1996. *The EM Algorithm and Extensions*. Wiley and Sons.

McLachlan, J. and D. Peel. 2000. *Finite Mixture Models*. New York, John Wiley and Sons Ltd.

Orphanides, A. and D. Zervos 1995. Rational addiction with learning and regret. *Journal of Political Economy* 103 4, 739–758.

Paolino, P. 2001. Maximum likelihood estimation of models with beta-distributed dependent variables. *Political Analysis* 9 (4), 325–346.

Peracchi, F. and V. Perotti. 2010. Subjective survival probabilities and life tables: Evidence from Europe. EIEF Working Paper 10/16.

Schafer, J. L. 1997. *Analysis of Incomplete Multivariate Data*, Volume 72. Monographs on Statistics ad Applied Probability. Chapman and Hall.

Schoenbaum, M. 1997. Do smokers understand the mortality effects of smoking? Evidence from the Health and Retirement Survey. *American Journal of Public Health* 87 (5), 755–759.

Smith, J. P. 1999. Healthy bodies and thick wallets: the dual relation between health and economic status. *The Journal of Economic Perspective* 13 (2), 145–166.

Smith, V. K., D. H. Taylor, and F. A. Sloan. 2001a. Longevity expectations and death: can people predict their own demise? *American Economic Review* 91, 1126–1134.

Smith, V. K., D. H. Taylor, F. A. Sloan, F. R. Johnson, and W. H. Desvousge. 2001b. Do smokers respond to health shocks? *The Review of Economics and Statistics* 83 (4), 675-687.

Smithson, M. and J. Verkuilen. 2006. A better lemon squeezer? Maximum likelihood regression with Beta distributed dependent variables. *Psychological Methods* 11 (1), 54–71.

Thaler, R. and H. Sheffrin. 1981. An economic theory of self-control. *Journal of Political Economy* 89 (2), 392–406.

van Doorslaer, E. and U.-G. Gerdtham. 2003. Does inequality in self-assessed health predict inequality in survival by income? Evidence from Swedish data. *Social Science and Medicine* 57, 1621–1629.

Viscusi, W. K. 1990. Do smokers underestimate risks? *Journal of Political Economy* 98 (6), 1253–1269.

Winston, G. 1980. Addiction and backsliding: The theory of compulsive consumption. *Journal of Economic Behaviour and Organization* 1, 295–234.

Wooldridge, J. 2002. Econometric analysis of cross section and panel data. Cambridge: MIT Press.

8 Tables and Figures

Table 1 - Target ages in the subjective survival probability question

Age	Target age	Target age - Age
50-65	75	20-25
56-60	75	15-19
61-65	75	10-14
66-70	80	10-14
71-75	85	10-14
76-80	90	10-14
81-85	95	10-14

Table 2 - Sample means in the overall sample and by subjective survival probability

N= 20524	overall sample	SSP < 50	SSP = 50	SSP > 50
subjective survival probability (SSP)	61.868	18.723	50	83.447
self-assessed health (SAH)	0.323	0.134	0.276	0.415
never smoker	0.512	0.530	0.519	0.502
current smoker	0.202	0.190	0.215	0.200
past smoker	0.286	0.279	0.266	0.298
years smoked (n=10509)	28.298	31.464	29.018	26.863
current smoker (n=4139)	36.539	39.852	36.261	35.498
past smoker (n=5876)	22.493	25.745	23.153	21.086
drinker	0.140	0.126	0.133	0.149
no physical exercise	0.083	0.170	0.079	0.053
obese	0.210	0.241	0.217	0.194
income	29185.080	23765.550	28613.970	31471.610
income 1	0.237	0.291	0.242	0.214
income 2	0.248	0.317	0.258	0.217
income 3	0.258	0.220	0.263	0.271
income 4	0.257	0.172	0.237	0.299
no education	0.046	0.063	0.045	0.040
less than Diploma	0.448	0.535	0.458	0.410
high school diploma	0.306	0.271	0.311	0.317
univeristy degree or higher	0.201	0.131	0.186	0.233
retired	0.474	0.641	0.468	0.414
employed	0.303	0.129	0.289	0.374
unemployed	0.034	0.025	0.036	0.035
sick	0.032	0.051	0.033	0.025
homemaker	0.155	0.150	0.171	0.150
married	0.750	0.661	0.738	0.789
separated or divorced	0.074	0.073	0.073	0.075
single	0.054	0.063	0.057	0.049
widowed	0.122	0.204	0.133	0.086
male	0.465	0.475	0.457	0.464
age	63.513	68.862	63.564	61.493
age 50-65	0.619	0.360	0.606	0.722
age 66-70	0.144	0.153	0.146	0.140
age 71-75	0.111	0.170	0.133	0.079
age 76-80	0.081	0.189	0.077	0.042
age 81-85	0.045	0.128	0.038	0.017
adl	1.830	2.062	1.669	1.631
iadl	1.770	2.035	1.661	1.522
gali	0.402	0.608	0.409	0.322
hospital admission in last year	0.125	0.203	0.126	0.095
tobacco related diseases	0.512	0.653	0.529	0.452
number of children	2.140	2.173	2.083	2.155
numeracy	3.400	3.124	3.424	3.492
deceased mother	0.749	0.875	0.764	0.695
age at death	74.986	74.150	74.393	75.679
deceased father	0.897	0.958	0.906	0.870
age at death	71.023	70.031	70.715	71.578
spouse age at death	63.062	66.097	62.692	60.646
northern	0.181	0.172	0.167	0.191
southern	0.272	0.237	0.278	0.281
central	0.547	0.590	0.555	0.527

Table 3 - Average SSP by SAH, smoking and socioeconomic status

		SSP				
	overall distribution	**= 75**	**= 80**	**= 85**	**= 90**	**= 95**
SAH						
excellent	76.633	78.607	73.918	67.056	62.055	59.742
	(0.477)	(0.493)	(1.548)	(2.137)	(3.686)	(7.341)
very good	69.766	72.970	67.039	61.772	50.135	45.795
	(0.357)	(0.380)	(1.046)	(1.446)	(1.964)	(2.722)
good	62.786	68.004	62.352	54.567	47.284	38.819
	(0.296)	(0.337)	(0.726)	(0.890)	(1.210)	(1.718)
fair	51.251	58.304	53.518	47.867	37.989	29.793
	(0.443)	(0.592)	(1.042)	(1.125)	(1.247)	(1.528)
poor	39.166	48.076	41.652	37.734	23.629	24.058
	(0.924)	(1.429)	(2.348)	(2.146)	(1.878)	(2.598)
smoking status						
never	61.034	68.527	61.391	53.123	42.537	33.378
	(0.276)	(0.316)	(0.666)	(0.771)	(0.944)	(1.244)
current	62.420	65.427	56.388	51.170	43.246	40.481
	(0.437)	(0.472)	(1.296)	(1.811)	(3.120)	(3.689)
past	62.969	70.291	61.841	53.205	42.047	36.345
	(0.377)	(0.416)	(0.956)	(1.137)	(1.357)	(1.959)
income quartile						
first	58.150	64.813	59.338	53.326	43.630	38.528
	(0.410)	(0.521)	(0.963)	(1.056)	(1.275)	(1.775)
second	57.780	65.037	59.817	51.931	41.527	34.380
	(0.414)	(0.529)	(0.918)	(1.046)	(1.326)	(1.686)
third	64.265	70.199	62.781	52.508	40.901	31.213
	0.380	*0.405*	*0.986*	*1.257*	*1.620*	*2.363*
forth	67.280	70.798	62.143	55.395	43.915	30.890
	(0.365)	(0.370)	(1.234)	(1.779)	(2.243)	(2.724)
education						
no education	55.983	63.446	57.207	55.418	42.014	43.898
	(0.985)	(1.553)	(2.174)	(1.941)	(2.562)	(3.963)
less than diploma	59.105	66.689	59.999	52.289	42.410	34.427
	(0.309)	(0.374)	(0.727)	(0.825)	(1.005)	(1.304)
high school diploma	63.516	68.277	62.672	51.663	41.130	34.233
	(0.347)	(0.379)	(0.907)	(1.231)	(1.579)	(2.111)
university degree or higher	66.865	71.188	61.271	56.287	45.034	32.839
	(0.400)	(0.410)	(1.245)	(1.633)	(2.144)	(3.143)

The table reports average SSP by SAH category, smoking status, income quartile and educational level. Averages are calculated for the overall sample , that is considering all target ages, and for each specific target age.

Table 4 - Distributions of self-assessed health and smoking variables

	SSP quartile				SAH				
	first	**second**	**third**	**fourth**	**excellent**	**very good**	**good**	**fair**	**poor**
SAH									
excellent	0.054	0.090	0.134	0.186					
	(0.002)	(0.008)	(0.005)	(0.005)					
very good	0.158	0.211	0.276	0.263					
	(0.004)	(0.012)	(0.006)	(0.006)					
good	0.396	0.462	0.406	0.387					
	(0.005)	(0.014)	(0.007)	(0.007)					
fair	0.295	0.206	0.158	0.135					
	(0.005)	(0.012)	(0.005)	(0.005)					
poor	0.097	0.030	0.026	0.030					
	(0.003)	(0.005)	(0.002)	(0.002)					
smoking status									
never	0.524	0.531	0.515	0.482	0.471	0.492	0.513	0.550	0.515
	(0.005)	(0.014)	(0.007)	(0.007)	(0.011)	(0.008)	(0.006)	(0.007)	(0.014)
current	0.204	0.199	0.192	0.207	0.201	0.221	0.202	0.183	0.198
	(0.004)	(0.012)	(0.006)	(0.006)	(0.009)	(0.006)	(0.004)	(0.006)	(0.011)
past	0.272	0.269	0.293	0.311	0.328	0.287	0.285	0.266	0.288
	(0.005)	(0.013)	(0.006)	(0.007)	(0.010)	(0.007)	(0.005)	(0.007)	(0.013)
years smoked									
current smok	37.759	36.899	35.064	35.578	35.166	34.880	36.782	37.948	39.332
	(0.285)	(0.657)	(0.361)	(0.354)	(0.534)	(0.366)	(0.283)	(0.441)	(0.817)
past smoker	24.344	22.803	20.591	21.198	18.678	20.462	21.886	26.320	28.806
	(0.279)	(0.717)	(0.335)	(0.319)	(0.449)	(0.355)	(0.268)	(0.404)	(0.768)

The “SSP quartile” panel describes the distributions of SAH and smoking status by SSP quartile and the average number of years smoked in each quartile; the “SAH” panel describes the distribution of smoking status and the average number of years smoked by SAH category.

Table 5 – Average SSP by target age and parental/spouse mortality

	SSP						SSP						SSP					
	overall distribution	**= 75**	**= 80**	**= 85**	**= 90**	**= 95**	**overall distribution**	**= 75**	**= 80**	**= 85**	**= 90**	**= 95**	**overall distribution**	**= 75**	**= 80**	**= 85**	**= 90**	**= 95**
	mother						**father**						**partner/spouse**					
dead	59.235	67.182	60.161	52.717	42.453	34.825	60.815	67.677	60.638	52.886	42.426	34.841	51.281	65.750	58.126	53.448	39.432	33.731
	(0.236)	(0.290)	(0.525)	(0.610)	(0.754)	(1.015)	(0.212)	(0.247)	(0.509)	(0.602)	(0.754)	(1.014)	(0.607)	(1.009)	(1.343)	(1.246)	(1.317)	(1.424)
alive	69.727	69.916	68.113	61.698	30.000	-	71.056	71.082	69.048	97.500	-	-	63.337	68.368	61.174	52.782	43.787	35.776
	(0.339)	(0.346)	(1.745)	(3.481)	(25.166)	-	(0.509)	(0.516)	(3.171)	(2.500)	-	-	(0.207)	(0.229)	(0.544)	(0.687)	(0.916)	(1.432)
age at death																		
< =50	56.159	65.035	58.494	51.957	38.241	34.988	60.086	67.694	59.194	51.879	44.402	36.821	60.736	67.781	61.618	52.941	35.971	41.923
	(0.937)	(1.223)	(2.198)	(2.325)	(2.525)	(3.366)	(0.694)	(0.864)	(1.475)	(1.837)	(2.425)	(3.424)	(1.375)	(1.656)	(3.377)	(3.894)	(4.529)	(4.403)
51 - 74	57.882	64.888	56.304	50.512	40.912	35.609	59.202	65.351	58.183	50.731	42.107	32.107	51.392	64.767	56.716	52.729	39.488	33.159
	(0.409)	(0.484)	(1.008)	(1.183)	(1.379)	(1.847)	(0.321)	(0.373)	(0.813)	(0.951)	(1.194)	(1.473)	(0.745)	(1.289)	(1.508)	(1.420)	(1.760)	(1.994)
75 - 79	59.623	67.639	57.589	50.680	42.417	30.945	61.471	68.966	60.678	52.219	40.635	38.542	42.652	70.000	67.500	59.595	37.316	32.022
	(0.586)	(0.672)	(1.436)	(1.584)	(2.107)	(2.621)	(0.523)	(0.593)	(1.313)	(1.505)	(1.752)	(2.621)	(1.913)	(7.508)	(5.846)	(4.251)	(2.749)	(3.123)
80 - 84	60.542	69.360	61.904	51.817	41.142	32.583	62.509	70.442	62.501	52.716	41.414	35.949	41.250	26.667	53.750	63.846	42.510	34.714
	(0.533)	(0.643)	(1.158)	(1.322)	(1.704)	(2.433)	(0.533)	(0.586)	(1.301)	(1.522)	(1.812)	(2.837)	(2.558)	(12.019)	(12.947)	(6.557)	(4.042)	(3.700)
85 - 89	60.112	69.287	63.893	54.191	41.248	27.797	63.265	71.330	64.820	56.152	40.725	33.948	46.744	-	-	43.333	53.308	36.471
	(0.629)	(0.803)	(1.175)	(1.567)	(1.906)	(2.622)	(0.689)	(0.770)	(1.533)	(1.864)	(2.554)	(3.150)	(5.384)	-	-	(11.450)	(8.729)	(8.366)
older than 90	61.255	70.441	64.121	58.427	51.049	44.116	64.216	70.960	66.538	63.659	50.752	38.653	31.333				35.000	28.889
	(0.707)	(1.073)	(1.358)	(1.526)	(1.995)	(2.542)	(0.880)	(1.142)	(1.754)	(2.057)	(2.974)	(3.803)	(7.096)				(8.851)	(10.599)

Table 6 - Information criteria, one class and two class model

	Single class model	Mixture model
-2logL	67602.406	54119.866
AIC	67812.406	54330.866
CAIC	68645.988	56215.959
BIC	68644.988	56214.959
no. parameters	105	211
N	20524	20524

The consistent Akaike information criterion (CAIC) is calculated as 2logL+[1+log(N)q] where q is the number of parameters.

Table 7 - Estimated coefficients from the finite mixture model

	survival expectations		subjective health		smoking duration		probability of starting	
variables	class 1	class2	class 1	class2	class 1	class2	class 1	class2
sah	0.319 (0.014) ***	0.430 (0.035) ***						
current smoker	-0.039 (0.124)	-0.121 (0.261)	0.148 (0.237)	0.264 (0.287)				
past smoker	0.153 (0.050) ***	0.027 (0.120)	0.148 (0.095)	0.277 (0.133) *				
ln(years smoked)*current smoker	-0.016 (0.034)	0.013 (0.072)	-0.061 (0.066)	-0.104 (0.080)				
ln(years smoked)*past smoker	-0.039 (0.016) **	0.008 (0.039)	-0.055 (0.032)	-0.106 (0.044) *				
income 2	-0.032 (0.019) *	-0.024 (0.046)	0.048 (0.038)	-0.088 (0.054)	-0.082 (0.038) *	-0.085 (0.047)	0.129 (0.034) ***	0.081 (0.045)
income 3	0.030 (0.020)	0.122 (0.048) **	0.009 (0.040)	0.103 (0.056)	-0.165 (0.038) ***	-0.136 (0.049) **	0.186 (0.035) ***	0.081 (0.048)
income 4	0.059 (0.021) **	0.041 (0.052)	0.099 (0.042) *	0.183 (0.059) **	-0.182 (0.041) ***	-0.141 (0.051) **	0.142 (0.037) ***	0.118 (0.050) *
less than Diploma	-0.085 (0.032) **	0.059 (0.078)	0.259 (0.077) ***	0.284 (0.106) **	0.167 (0.068) *	0.219 (0.080) **	0.252 (0.060) ***	0.171 (0.077) *
high school diploma	-0.056 (0.034)	0.117 (0.082)	0.420 (0.078) ***	0.409 (0.108) ***	0.158 (0.070) *	0.201 (0.083) *	0.248 (0.062) ***	0.224 (0.081) **
univeristy degree or higher	0.018 (0.036)	0.068 (0.088)	0.653 (0.080) ***	0.537 (0.112) ***	0.020 (0.071)	0.029 (0.086)	0.209 (0.065) **	0.157 (0.085)
retired	0.019 (0.019)	0.086 (0.047) *	-0.034 (0.038)	0.047 (0.055)	-0.109 (0.041) **	-0.017 (0.052)		
employed	0.097 (0.021) ***	0.235 (0.051) ***	0.203 (0.039) ***	0.233 (0.056) ***	-0.188 (0.040) ***	-0.055 (0.053)		
male	-0.087 (0.015) ***	-0.115 (0.036) ***	0.022 (0.028)	0.051 (0.040)	0.008 (0.027)	0.053 (0.035)	0.833 (0.023) ***	0.820 (0.032) ***
age 50-65	0.908 (0.037) ***	1.461 (0.080) ***	0.205 (0.084) *	0.357 (0.106) ***	0.006 (0.067)	0.187 (0.074) *	0.382 (0.062) ***	0.421 (0.071) ***
age 66-70	0.752 (0.038) ***	1.165 (0.082) ***	0.113 (0.086)	0.229 (0.109)	0.031 (0.068)	0.068 (0.076)	0.154 (0.067) *	0.115 (0.079)
age 71-75	0.498 (0.038) ***	0.884 (0.084) ***	0.022 (0.088)	0.077 (0.113)	-0.024 (0.069)	0.069 (0.077)	0.072 (0.069)	0.153 (0.082)
age 76-80	0.195 (0.040) ***	0.333 (0.085) ***	0.014 (0.094)	-0.104 (0.120)	-0.151 (0.072) *	0.027 (0.078)	0.002 (0.073)	0.072 (0.084)
married	0.066 (0.028) **	0.342 (0.075) ***	0.010 (0.054)	0.303 (0.088) ***	-0.172 (0.058) **	-0.232 (0.079) *		
separated or divorced	0.071 (0.035) *	0.084 (0.092)	-0.084 (0.069)	0.470 (0.106) ***	0.057 (0.071)	0.081 (0.096)		
widowed	-0.064 (0.053)	0.235 (0.130) *	-0.031 (0.067)	0.360 (0.104) ***	-0.011 (0.071)	-0.040 (0.094)		
target age - age	-0.015 (0.002) ***	-0.024 (0.005) ***						
partner age at death < 50	0.073 (0.063)	0.119 (0.156)						
partner age at death 51-74	0.029 (0.050)	-0.073 (0.118)						
mother died	-0.100 (0.019) ***	-0.054 (0.046)						
mother age at death < 50	-0.012 (0.030)	-0.113 (0.072)						
mother age at death 75-79	0.007 (0.021)	-0.071 (0.053)						
mother age at death 80-84	0.038 (0.020) *	0.001 (0.049)						
mother age at death 85-89	0.040 (0.022) *	-0.005 (0.054)						
father died	-0.141 (0.024) ***	-0.139 (0.058) ***						
father age at death < 50	0.038 (0.024)	0.048 (0.057)						
father age at death 75-79	0.085 (0.019) ***	0.012 (0.047)						
father age at death 80-84	0.108 (0.019) ***	0.089 (0.048) *						
father age at death 85-89	0.161 (0.024) ***	0.058 (0.058)						
numeracy	0.002 (0.007)	0.050 (0.016) ***						
adl			-0.155 (0.048) **	-0.516 (0.095) ***				
iadl			-0.176 (0.035) ***	-0.261 (0.052) ***				
gali			-1.049 (0.030) ***	-0.986 (0.042) ***				
hospital admission in last year			-0.246 (0.045) ***	-0.299 (0.063) ***				
drinker					0.132 (0.030) ***	0.100 (0.039) *		
no physical exercise					0.265 (0.054) ***	0.286 (0.058) ***		
obese					-0.107 (0.029) ***	-0.127 (0.037) ***		
tobacco-related diseases					-0.128 (0.024) ***	-0.039 (0.031)		
number of children					-0.006 (0.009)	0.046 (0.011)		
northern	-0.132 (0.022) ***	-0.263 (0.050) ***	0.636 (0.041) ***	0.707 (0.056) ***	-0.096 (0.039) *	-0.123 (0.048) *	0.280 (0.037) ***	0.465 (0.049) ***
central	-0.106 (0.016) ***	-0.330 (0.040) ***	0.011 (0.032)	0.111 (0.045) *	-0.164 (0.032) ***	-0.165 (0.041) ***	0.079 (0.029) **	0.134 (0.039) ***
intercept	-0.059 (0.071)	-0.542 (0.171)	-0.820 (0.124) ***	-1.343 (0.173) ***	4.181 (0.114) ***	3.764 (0.140) ***	-1.174 (0.084) ***	-1.046 (0.102) ***
ϕ	6.735 (0.077)	0.421 (0.006)			1.422 (0.021) ***	1.454 (0.028) ***		
N	20524		logL	-27059.933				

Standard errors in parenthesis. Level of significance: *** 1%; ** 5%; * 10%.

Table 8 - Estimated odds ratios for SSP and average partial effects for SAH

	survival expectation		subjective health	
variables	class 1	class2	class 1	class2
sah	1.376	1.537		
current smoker	0.962	0.886	0.042	0.074
past smoker	1.166	1.028	0.042	0.077
ln(years smoked)*current smoker	0.984	1.013	-0.018	-0.028
ln(years smoked)*past smoker	0.962	1.008	-0.016	-0.029
income 2	0.969	0.977	0.014	-0.024
income 3	1.030	1.130	0.003	0.029
income 4	1.060	1.042	0.028	0.053
less than Diploma	0.919	1.061	0.068	0.074
high school diploma	0.946	1.125	0.114	0.109
univeristy degree or higher	1.018	1.070	0.185	0.147
retired	1.020	1.089	-0.010	0.013
employed	1.101	1.265	0.060	0.067
male	0.916	0.892	0.006	0.014
age 50-65	2.479	4.312	0.058	0.099
age 66-70	2.121	3.207	0.031	0.062
age 71-75	1.646	2.420	0.006	0.020
age 76-80	1.215	1.396	0.004	-0.026
married	1.068	1.407	0.003	0.080
separated or divorced	1.074	1.088	-0.024	0.128
widowed	0.938	1.265	-0.009	0.097
target age - age	0.985	0.976		
partner age at death < 50	1.076	1.127		
partner age at death 51-74	1.029	0.930		
mother died	0.905	0.948		
mother age at death < 50	0.988	0.893		
mother age at death 75-79	1.008	0.931		
mother age at death 80-84	1.039	1.001		
mother age at death 85-89	1.040	0.995		
father died	0.869	0.870		
father age at death < 50	1.039	1.049		
father age at death 75-79	1.089	1.012		
father age at death 80-84	1.115	1.093		
father age at death 85-89	1.175	1.060		
numeracy	1.002	1.051		
adl			-0.045	-0.137
iadl			-0.051	-0.071
gali			-0.300	-0.276
hospital admission in last year			-0.068	-0.081
drinker				
no physical exercise				
obese				
tobacco-related diseases				
number of children				
northern	0.877	0.769	0.193	0.208
central	0.900	0.719	0.003	0.031
baseline individual	0.754	0.475		

Table 9 - Predicted average SSP and average probability of SAH=1 for hypothetical scenarios and by target age

	SSP											
	overall		**= 75**		**= 80**		**= 85**		**= 90**		**= 95**	
	class 1	class 2	class 1	class 2	class 1	class 2	class 1	class 2	class 1	class 2	class 1	class 2
Predicted average SSP												
observed smoking bahaviour	0.581	0.652	0.622	0.713	0.585	0.650	0.523	0.585	0.448	0.452	0.401	0.373
smoking scenario	0.564	0.630	0.604	0.691	0.567	0.626	0.505	0.560	0.430	0.426	0.384	0.349
quitting scenario	0.609	0.661	0.649	0.721	0.613	0.660	0.552	0.595	0.477	0.462	0.429	0.383
smoking free scenario	0.573	0.656	0.614	0.716	0.576	0.654	0.514	0.589	0.440	0.456	0.393	0.377
Predicted average probability of SAH=1												
observed smoking bahaviour	0.322	0.326	0.336	0.349	0.309	0.312	0.284	0.270	0.282	0.224	0.278	0.250
smoking scenario	0.343	0.362	0.358	0.387	0.331	0.349	0.305	0.306	0.302	0.257	0.298	0.285
quitting scenario	0.344	0.366	0.358	0.391	0.331	0.353	0.305	0.309	0.303	0.261	0.299	0.288
smoking free scenario	0.301	0.289	0.315	0.311	0.289	0.275	0.264	0.236	0.262	0.194	0.258	0.217

Table 10 – Sample means in the heterogenous population

variables	sample 1	sample 2
	N=15093	N=5431
subjective survival probability (SSP)	58.974	69.908
self-assessed health (SAH)	0.324	0.321
never smoker	0.524	0.478
current smoker	0.196	0.217
past smoker	0.280	0.305
years smoked (n=10509)	28.062	28.897
current smoker (n=4139)	36.295	37.150
past smoker (n=5876)	22.290	23.013
drinker	0.138	0.147
no physical exercise	0.071	0.119
obese	0.206	0.220
income 1	0.234	0.245
income 2	0.242	0.262
income 3	0.263	0.247
income 4	0.261	0.246
no education	0.045	0.048
less than Diploma	0.437	0.476
high school diploma	0.309	0.298
univeristy degree or higher	0.209	0.178
retired	0.463	0.504
employed	0.310	0.284
unemployed	0.035	0.031
sick	0.029	0.040
homemaker	0.161	0.138
married	0.749	0.753
separated or divorced	0.074	0.074
single	0.057	0.045
widowed	0.120	0.127
male	0.463	0.469
age	63.234	64.289
age 50-65	0.628	0.594
age 66-70	0.148	0.133
age 71-75	0.109	0.116
age 76-80	0.076	0.096
age 81-85	0.039	0.061
adl	0.128	0.200
iadl	0.217	0.323
gali	0.395	0.423
tobacco related diseases	0.509	0.520
hospital admission in last year	0.119	0.140
number of children	2.112	2.220
numeracy	3.440	3.288
deceased mother	0.744	0.763
age at death	74.993	74.967
deceased father	0.895	0.905
age at death	71.114	70.772
spouse age at death	62.794	63.766
northern	0.168	0.218
southern	0.269	0.279
central	0.563	0.502

Table 11 - Probit model for class membership, standard errors in brackets

variables	Probability of class membership	
income 2	-0.029	(0.028)
income 3	0.034	(0.030)
income 4	0.004	(0.032)
less than Diploma	-0.069	(0.047)
high school diploma	-0.063	(0.051)
univeristy degree or higher	-0.004	(0.054)
retired	-0.088	(0.029)
employed	-0.042	(0.032)
male	-0.021	(0.021)
age 50-65	0.163	(0.052)
age 66-70	0.269	(0.053)
age 71-75	0.184	(0.053)
age 76-80	0.105	(0.054)
married	-0.122	(0.046)
separated or divorced	-0.115	(0.056)
widowed	0.002	(0.081)
target age - age	0.002	(0.003)
partner age at death < 50	-0.002	(0.095)
partner age at death 51-74	-0.096	(0.073)
mother died	0.001	(0.028)
mother age at death < 50	-0.008	(0.045)
mother age at death 75-79	0.023	(0.032)
mother age at death 80-84	0.013	(0.030)
mother age at death 85-89	-0.017	(0.033)
father died	-0.008	(0.036)
father age at death < 50	-0.051	(0.035)
father age at death 75-79	-0.003	(0.029)
father age at death 80-84	0.029	(0.029)
father age at death 85-89	0.001	(0.036)
numeracy	0.057	(0.010)
adl	-0.024	(0.019)
iadl	-0.020	(0.016)
gali	0.028	(0.021)
hospital admission in last year	-0.060	(0.029)
drinker	-0.078	(0.028)
no physical exercise	-0.270	(0.037)
obese	-0.034	(0.024)
tobacco-related diseases	0.023	(0.020)
number of children	-0.020	(0.007)
northern	-0.199	(0.031)
central	0.049	(0.024)
intercept	0.564	(0.109)

Figure 1 – Distribution of SSP, average SSP and Life Tables probability by target age and European regions