Survival expectations, subjective health and smoking: evidence from SHARE

Silvia Balia

Abstract This article aims to assess how the risk perceptions of smokers affect survival expectations and subjective health. Data from the Survey of Health, Ageing and Retirement in Europe, which include a numerical measure of subjective survival probability, are used to estimate a joint recursive system of equations that describe the relationships among survival expectations, subjective health status and smoking duration. A finite mixture model is used to address endogeneity and unobservable heterogeneity. This approach allows for two types of individuals with different observable characteristics to be identified in the examined population. We find that only in the population of the first type, current and former smokers incorporate the effects of smoking duration into their assessments of survival probabilities. For both types, quitting smoking affects current perceptions of smoking risks, causing the overestimation of both survival probability and subjective health.

S. Balia (✉)
Department of Economics and Business, University of Cagliari and CRENoS,
Viale S. Ignazio, 78, Cagliari 09123, Italy
e-mail: silvia.balia@unica.it

1 Introduction

Measures of beliefs about survival probability are frequently available in individual survey data. These measures are of potential interest because they may be used in models of health investments that examine individuals' choices of health-related behaviours over the life cycle. However, for each individual in the population, a self-assessment of survival probability might be largely influenced by the individual's perceptions of certain health risks. These perceptions depend on how individuals evaluate the costs and benefits of their behaviours to their current health and their future mortality risk. Perceptions of health risks might influence individuals' incentives to adopt healthy lifestyles. This paper contributes to the extant literature on smoking and risk perception that addresses the notion of information constraints that may cause the costs and benefits associated with health-related behaviours to be estimated erroneously (see Cawley and Ruhm 2011). The existing empirical studies that link smoking and individuals' subjective perception of their health status and mortality risk have largely been based on US data and have generated mixed results.[1] In this work, we use European data to examine how individuals' perceptions of the health effects of smoking influence their survival expectations and subjective health.

Over the past century, American and European populations have experienced positive trends in longevity, although the former is characterised by lower life expectancy. This gap largely reflects the higher prevalence of non-communicable diseases, the main causes of premature mortality in elderly populations, in the US than in Europe; in the middle-aged American population, these diseases are frequently related to various risk factors such as smoking and obesity (Michaud et al. 2011). However, the most recent estimations of daily tobacco consumption reveal that, in 2008, the estimated prevalence of smoking is about 15.6 % in the US but much higher in Europe, where the highest prevalences of smoking are observed in Greece and Austria (44.3 and 39.8 %, respectively) and the lowest are found in Italy and Belgium (19.6 and 19.8). Cross-country variation in smoking can be attributed in part to differences in beliefs among countries with respect to the health consequences of smoking (Cutler and Glaeser 2009). During the past decade, many European countries have exhibited a commitment to both tobacco control action plans and smoke-free legislative initiatives. A failure to fully consider individual perceptions of the consequences of current and former smoking habit could account for the low levels of success of these interventions. This provides a motive for studying perception of the health and mortality risks that are associated with smoking in Europe.

---

[1] Schoenbaum (1997) finds that non-smokers overestimate the mortality risk of smoking while heavy smokers underestimate the negative effect of smoking intensity on survival probability. Viscusi (1990) and Viscusi and Hakes (2008) find that smokers overestimate the smoking-related risks of lung cancer, the life expectancy loss and total mortality risk. They also show that higher risk perception lowers the probability of beginning to smoke and increase the probability of quitting. Smith et al. (2001a,b) find significant heterogeneity among smokers. Heavy smokers are excessively optimistic about their future survival, whereas current smokers reduce their survival expectations more dramatically than former and never smokers when they experience smoking-related health shocks. More evidence supporting underestimation of the health risks of smoking is provided by Hammar and Johansson-Stenman (2004) and Khwaja et al. (2007).

Furthermore, it has been well documented that smoking may produce immediate side effects, such as increases in pulse rate, blood pressure and weight, and that mortality risks are greater for smokers than for never smokers because the probability of the onset of various health issues increases as a result of the prolonged consumption of tobacco. Therefore, in this work, we seek to assess the risk perception of current, former and never smokers and to examine perceptions of the short- and long-term health effects of smoking, estimating a model for individual survival expectations, subjective health and smoking habit. Our approach allows us to understand whether individuals believe that the detrimental short- and long-term effects of smoking are reversible. A belief in the reversibility of these effects would suggest that the true health-related effects of smoking would be underestimated for individuals who eventually quit smoking.[2] This issue is particularly concerning because anti-tobacco campaigns often disseminate the information that smoking is bad for your health but that quitting cancels out the long-term risks.

Our analysis of heterogeneity in risk perception among smokers must be interpreted in the context of the myopic and rational models of smoking behaviour, which derive from the traditional economic approach and treat smoking as an addictive good.[3] Myopic and rational models have different implications. Rational individuals smoke only if the benefits of smoking outweigh the costs of smoking, whereas present-oriented (myopic) individuals may be more addicted to smoking than rational individuals. Furthermore, rational addicts smoke more if they expect future prices to fall, whereas myopic addicts do not adjust their behaviours in this manner. Empirical results strongly reject the hypothesis of myopic behaviour and support the model of rational addiction (Becker et al. 1994; Chaloupka 1991). Assuming that the myopic model can simply be nested within the rational model and controlling for unobservable heterogeneity, Arcidiacono et al. (2007) find that both models predict that smoking rates decrease with age; as individuals become older, their health worsens, illnesses occur more frequently, and smoking becomes less attractive. The rational model predicts the occurrence of a sharp decline in smoking rates after the age of 62 that is followed, however, by a significant upward trend in smoking behaviour after a cut-off age of 80. This end-of-life effect can be defined as a 'rationally myopic' attitude of older individuals, who expect to die soon and are, therefore, less concerned about the future effects of smoking.

First, in this paper, we specify a simultaneous recursive model for survival expectations, subjective health and smoking behaviour conditional on observed individuals

---

[2] Carbone et al. (2005) compare and contrast two hypothetical and extreme scenarios: in one scenario, individuals' beliefs about the probability of dying depend on both their current health and prior smoking habit (the *irreversible* case), whereas in the other scenario, these beliefs depend only on health status (the *reversible* case).

[3] Myopic models describe smokers as irrational addicts who fail to consider the impact of current consumption on their future consumption and utility (see Thaler and Shefrin 1981; Winston 1980). Tobacco consumption only depends on current price and past consumption; individuals care more about current utility derived from smoking than about the future. The negative effects of smoking on an individual's future health and survival are not considered. The theory of rational addiction regards addictive tobacco consumption as a rational choice (Becker and Murphy 1988; Becker et al. 1994). Smokers are forward-looking individuals who care about the effects of today's consumption on their future consumption and utility; the detrimental effects of smoking are internalised when making consumption decisions.

characteristics including valuable information about optimism. We propose a finite mixture approach to account for unobservable factors that might simultaneously influence subjective assessments and smoking, thereby addressing the issues of reverse causation and endogeneity in the relationship between smoking and individuals' beliefs. Second, we use data on elderly Europeans who responded to the Survey of Health, Ageing and Retirement in Europe (SHARE) and for whom unique information about survival expectations at different target ages has become available since the SHARE data have been released. By contrast, most of the existing evidence about the risk perception of smokers is based on data from the US. To account for heterogeneity in the risk perceptions of respondents, various types of smokers are differentiated by smoking status at the time of the survey and by the duration of smoking habit over the life cycle; this approach defines the appropriate scope for analyses of the hazard of quitting smoking.

We identify two classes, or types, of individuals in the population; these types differ with respect to both observed and unobserved characteristics. On average, compared with the second type, the first type of individuals has higher survival expectations and lower subjective health; it includes a smaller proportion of smokers, and these smokers have smoked for a shorter duration. For the individuals of the first type, smoking is an important predictor of survival expectations, and smokers incorporate the (long-term) risks of smoking duration into their assessments of survival probabilities. However, for both current and former smokers, this effect vanishes as the duration of smoking habit increases. One important result, attributable to present-oriented (myopic) behaviour, is that for both classes, former smokers appear to perceive the harmful consequences of smoking as reversible and, particularly in the second class, overestimate both survival probability and health status. Subjective health assessments are significantly less positive for current smokers than for other survey respondents in both classes.

2 Data and variables

We use data from the first wave (2004) of the SHARE, which is designed in accordance with the approaches of the Health and Retirement Study (HRS) and the English Longitudinal Study of Ageing (ELSA).[4] The target population of this survey is non-institutionalised individuals aged 50 and older. Spouses are also interviewed. The

---

[4] The SHARE collects data in 2006–2007 (wave 2) and in 2008–2009 (SHARELIFE). Ideally, the longitudinal dimension should be exploited in the analysis. However, we cannot used SHARELIFE because it does not investigate survival expectations and smoking behaviour. Panel data analysis is excellent for addressing unobservable heterogeneity and the endogeneity of regressors (e.g. internal instruments could be used in the absence of valid instrumental variables). However, the availability of only two survey waves limits the potential for using panel analysis (only a balanced sample can be used but at the cost of a loss of about 45 % of observations due to attrition and item non-response). Attrition bias represents another challenge to the use of panel analysis because only the respondents who are less frail would remain in the sample. Moreover, in addition to mortality, which is *per se* related to the health measures that are utilised in this study, other reasons for drop-outs might create bias in a panel analysis (e.g. individuals whose health dramatically deteriorates between waves may not participate in wave 2 of the survey due to either a serious illness or institutionalisation). Thus, appropriate techniques that are based on inverse probability weights, or calibrated weights, should be employed for a panel analysis. However, this solution is not viable because the EM algorithm used in the present study already utilises weights in the estimation procedure.

SHARE provides rich information about not only respondents' health and lifestyles but also their survival expectations; previous European surveys have not collected survival expectation information. The complete sample consists of 31,115 individuals and features a response rate of about 85 %. For the purpose of our study and because of item non-response in the variables of interest, the sample used in the analysis consists of 20,285 respondents, who are between 50 and 85 years of age, from northern (Denmark and Sweden), central (Austria, Belgium, Germany, France and the Netherlands) and southern Europe (Italy, Spain and Greece).[5]

2.1 The measures of survival expectations, subjective health and smoking

Survival expectations are measured by a numerical indicator of subjective survival probability (SSP), derived from responses to the following question: 'What are the chances that you will live to be age T or more?'. Responses are driven by a card that reports a sequence of numbers from 0 ('absolutely no chance') to 100 ('absolutely certain'). SSP is, therefore, a continuous random variable, bounded between 0 and 100. The different target ages (T) depend on the age category of each respondent and reflects the fact that life table survival rates do not decline monotonically but increase within each age class. In this paper we consider target ages of 75, 80, 85, 90 and 95.[6]

The SSP question is the ninth out of eleven questions about the predicted probabilities of future events.[7] Although a warm-up question is asked to help respondents feel at ease with the notion of probabilities, we further increase the reliability of the SSP responses by excluding individuals who provided non-coherent responses to two questions about the probability that standards of living will be better or worse in the future. These individuals are 'hidden outliers' who could bias our estimates because they most likely provide unreliable subjective assessments of any future events.[8]

The elicitation of survival expectations through the use of probabilities is typically preferred to the alternative of utilising qualitative responses. Probabilities allow for a better comparison across individuals than qualitative responses; in addition, qualitative responses (such as 'likely' or 'very likely') may vary based on cognitive, linguistic and

---

[5] Individuals older than 85 years of age have been excluded because they are not heavily represented in the target sample and because they may have provided assessments that were less accurate than the responses of other survey participants. Moreover, most individuals of the same age are likely to be living in nursing homes, and the SHARE is not representative of the institutionalised population.

[6] These target ages correspond to the age categories 50–65, 66–70, 71–75, 76–80 and 81–85, with a difference from each respondent's current age ranging between 14 and 25 years. We exclude from the individuals those who were not matched to the appropriate target age.

[7] Guiso et al. (2005) and Delavande and Rohwedder (2011) show that the SHARE indicator of SSP exhibits similar characteristics to the SSP indicator in other major surveys, such as HRS and ELSA.

[8] The two events living better and living worse in the future are complements; therefore, the two subjective probabilities associated with these questions must sum to one. A subjective probability of 0 for both events indicates a high expected probability that the standard of living in the future will be unchanged. Using a tolerance level of 0.10, we exclude 831 individuals (about 3 % of the original sample) for whom the sum of the probabilities for the two questions is greater than 1.10. As a robustness check, we estimate the simultaneous recursive model with and without the use of this sample restriction. Estimated coefficients are very similar in magnitude; the signs and significance levels are unaffected. The log-likelihood and the information criteria favour the model estimated using the sample that excludes non-coherent responses.

cultural differences and typically suffer from response bias. Furthermore, the internal consistency of probabilities can be assessed (Dominitz and Manski 1997; Manski 2004). Another advantage of using a quantitative measure of survival expectations is that it is comparable with both observed mortality data and probabilities computed from life tables. We estimate a probit model for the probability of dying between waves of the SHARE to examine how average SSP vary between survivors and deceased individuals in our sample.[9] Results, reported in Table B.1 (available in the electronic supplementary material), suggest that survivors report higher SSP (of about 63) than individuals who die between waves (about 44) and confirm that individuals with higher SSP are more likely to survive to the next wave.[10] SSP is generally considered to be a better predictor of future mortality than objective life table hazard rates (Peracchi and Perotti 2010; Hurd and McGarry 1995; Hurd et al. 1999; Hurd and McGarry 2002). We compare SSP with the Human Mortality Database period life tables for 2004 and find that the average SSP is lower than the average survival probability calculated from life tables at ages 75, 80 and 85 but higher than that calculated at ages 90 and 95 (see Fig. B.1 in the electronic supplementary material). This might capture the fact that SSP most likely depends on both observable and unobservable individual characteristics, not included in life tables, that influence beliefs (such as an individual's level of optimism).

The primary disadvantage of using SSP is that heaping occurs at certain focal values, such as 0, 50, 100 and values that end with a zero, as shown in Fig. B.1. About 4.1 % of respondents report having no chance to survive to the target age; 25.2 % report an SSP equal to 50, and 15.8 % of respondents claim that they are certain to survive to the target age. Responses are more concentrated at high values (60, 70, 80 and 90). We address this issue in the econometric modelling by selecting an appropriate parametric distribution for SSP.

Survival beliefs are strictly related to individuals' perception of their own health. In the SHARE questionnaire, respondents are asked 'How is your health?' and can answer 'excellent', 'very good', 'good', 'fair' or 'poor'. These categorical responses are assumed to correspond to a continuous latent variable that measures subjective (or perceived) health. The indicator of self-assessed health (SAH) that can be derived from this question has commonly been used as a measure of general health status (see Deaton and Paxson 1998) and is known to be both a good indicator of morbidity and a powerful predictor of future health and mortality (Doorslaer and Gerdtham 2003). The recent literature has focused on the issue of reporting heterogeneity in SAH, which should be accounted for in measurements of health-related socioeconomic inequalities.

---

[9] Winter (2008) use information about the vital status of respondents at subsequent waves to demonstrate that, for each level of SAH, the average SSP is higher for survivors than for individuals who are deceased by wave 2 of the SHARE; he finds that SSP reported in wave 1 well predicts mortality that occurs between waves.

[10] Note that information on vital status at wave 2 is unavailable for about 29 % of our sample; therefore, statistics and estimates are computed from a smaller sample. As a robustness check, we have estimated a probit model for the probability of dying between waves 2 and 3. Results, available upon request, confirm that SSP is a good predictor of subsequent mortality.

In this work, we use a binary version of SAH that takes a value of 1 if reported health is excellent, very good or good and 0 otherwise.[11]

Information on smoking habit in the SHARE mainly derives from the questions 'Have you ever smoked cigarettes, cigars, cigarillos or a pipe daily for a period of at least one year?' and 'Do you smoke at the present time?', which allow us to build three binary indicators that indicate whether respondents have never smoked, are current smokers or are former smokers at the time of the interview. The question 'For how many years did you smoke?' provides the required information to construct a duration time variable that indicates the number of years that each respondent has spent smoking. This variable, which is right-censored at the time of the interview for current smokers (i.e. complete spells of smoking are observed only for former smokers), provides us with the appropriate scope to analyse the hazard of quitting.

2.2 Descriptive statistics

As reported in Table 1, where the variables used in the analysis are defined, the average SSP is 62 %; about 72.5 % of respondents report that they are in excellent, very good or good health. Current smokers, who have been smoking for an average of 36.6 years, comprise 20 % of the sample. About 28.8 % of smokers, who had smoked for an average of 22.5 years, have quit by the time of the interview.

Since a 50 % SSP might reflect 'epistemic uncertainty' rather than probabilistic thinking (Bruine de Bruin et al. 2002), or it could represent a genuine survival expectation (Hill et al. 2005), Table 1 also compares individuals with SSP values that are equal to 50 % with individuals who report lower and higher probabilities. The proportion of respondents in better health increases monotonically as we move from the sub-sample with SSPs lower than 50 to the sub-sample with higher SSP values. The same trend is found for disability measures (i.e. gali, adl and iadl). Moreover, as we move from the sub-sample with SSPs lower than 50 to the sub-samples with higher SSPs, the proportions of never smokers, sedentary individuals and obese individuals decrease, whereas the proportion of drinkers increases. However, current smokers are more concentrated in the sub-samples with SSPs that are equal to or greater than 50 %, whereas former smokers are concentrated in the sub-sample with the highest SSPs. This figures might reflect a form of cognitive dissonance that causes smokers to underestimate the negative effects of smoking on their survival (Chapman et al. 1983). As expected, longer smoking durations are associated with lower survival expectations for both current and former smokers. Individuals with poor socioeconomic statuses are concentrated in the sub-sample of individuals who report SSPs lower than 50. The

---

[11] Etilè and Milcent (2006) suggest the aggregation of responses into two broader categories as a method of reducing the reporting heterogeneity bias. In particular, they transform SAH into a binary indicator of poor/non-poor health. To account for reporting style Jürges (2007) proposes a five-point scale. As a robustness check, we re-estimated the model presented in Sect. 4 with ordinal values of SAH. Coefficients in the SSP equation of both models are very similar in magnitude; have the same signs and significance levels. For the purpose of our analysis, the binary version of SAH appears to accurately capture both the correlation between subjective health and survival expectations and the heterogeneity among different types of smokers with respect to SAH. Results are available upon request.

Table 1 Sample means and variable definitions

Variable name	Variable definition	Overall sample $N = 20{,}285$	SSP < 50 $N = 4{,}105$	SSP = 50 $N = 5{,}096$	SSP > 50 $N = 11{,}084$
SSP	Subjective survival probability	62.0	18.7	50.0	83.4
SAH	1 If self-assessed health is excellent, very good or good, 0 otherwise	0.725	0.488	0.709	0.820
Never smoker	1 If never smoked in her life, 0 otherwise	0.510	0.527	0.516	0.501
Current smoker	1 If smokes at the time of the interview, 0 otherwise	0.203	0.191	0.217	0.200
Smoking duration	Number of years spent smoking by current smokers	36.555	39.852	36.261	35.498
Former smoker	1 If quit smoking before the interview, 0 otherwise	0.288	0.282	0.267	0.299
Smoking duration	Number of years spent smoking by former smokers	22.483	25.745	23.153	21.086
Drinker	1 If drinks more than 2 glasses of alcohol almost every day, 0 otherwise	0.141	0.127	0.133	0.150
No physical exercise	1 If never or almost never engages in neither moderate nor vigorous physical activity, 0 otherwise	0.082	0.168	0.078	0.053
Obese	1 If body mass index <30 (28.6) for men (women), 0 otherwise	0.212	0.245	0.220	0.196
1st Income quartile	1 If bottom income quartile, 0 otherwise	0.235	0.287	0.239	0.213
2nd Income quartile	1 If second income quartile, 0 otherwise	0.248	0.319	0.259	0.217
3rd Income quartile	1 If third income quartile, 0 otherwise	0.259	0.221	0.264	0.271
4th Income quartile	1 If top income quartile, 0 otherwise	0.258	0.172	0.238	0.299
No education	1 If has no educational qualifications, 0 otherwise	0.045	0.062	0.044	0.039
Less than Diploma	1 If has less than high school diploma, 0 otherwise	0.446	0.533	0.456	0.409
High school diploma	1 If has high school diploma, 0 otherwise	0.307	0.272	0.313	0.318
Univeristy degree	1 If has university degree or higher qualitifications, 0 otherwise	0.202	0.133	0.187	0.234
Retired	1 If retired, 0 otherwise	0.474	0.643	0.468	0.414
Employed	1 If employed, 0 otherwise	0.305	0.130	0.292	0.375
Unemployed	1 If unemployed, 0 otherwise	0.034	0.026	0.036	0.035

Table 1 continued

Variable name	Variable definition	Overall sample N = 20,285	SSP < 50 N = 4,105	SSP = 50 N = 5,096	SSP > 50 N = 11,084
Sick	1 If absent from work due to sickness, 0 otherwise	0.032	0.051	0.032	0.025
Homemaker	1 If housekeeper, 0 otherwise	0.153	0.148	0.168	0.148
Married	1 If married, 0 otherwise	0.752	0.664	0.740	0.791
Separated or divorced	1 If separated or divorced 0 otherwise	0.074	0.072	0.073	0.075
Single	1 If single, 0 otherwise	0.053	0.061	0.056	0.049
Widowed	1 If widowed, 0 otherwise	0.121	0.203	0.132	0.085
Age	Age in years	63.471	68.830	63.508	61.468
Adl	Number of limitations with activities of daily living	0.133	0.323	0.117	0.069
Iadl	Number of limitations with instrumental activities of daily living	0.099	0.259	0.081	0.047
Gali	1 If limited in usual activities, 0 otherwise	0.400	0.607	0.406	0.321
Hospital admission	1 If during the last twelve months had at least One hospital admission, 0 otherwise	0.125	0.202	0.126	0.096
Tobacco-related diseases	1 Has been diagnosed a tobacco-related disease, 0 otherwise	0.511	0.654	0.528	0.450
Number of children	Number of children living with parents	0.624	0.474	0.608	0.686
Numeracy	Numeracy score	3.410	3.139	3.434	3.500
Pessimism	1 If the individual is pessimistic, 0 otherwise	0.126	0.219	0.137	0.088
Deceased mother	1 If mother is already dead, 0 otherwise	0.749	0.875	0.764	0.695
Mother age at death	Age of death of mother	75.008	74.166	74.412	75.701
Deceased father	1 If father is already dead, 0 otherwise	0.897	0.958	0.906	0.870
Father age at death	Age of death of father	71.018	70.024	70.722	71.566
Partner age at death	Age of death of partner/spouse	63.039	66.034	62.655	60.662
Northern Europe	1 If from Denmark or Sweden, 0 otherwise	0.182	0.173	0.168	0.191
Southern Europe	1 If from Italy, Spain or Greece, 0 otherwise	0.270	0.236	0.276	0.279
Central Europe	1 If from Austria, Belgium, Germany, France or Netherlands, 0 otherwise	0.548	0.591	0.556	0.529

latter individuals have experienced a higher proportion of deaths of their mothers, fathers and spouses than individuals with higher expectations. However, individuals who report higher SSP values had spouses who died at a younger age. Overall, the

observable differences across these three sub-groups suggest that a response of a 50 % SSP is likely to reflect a genuine answer.

The observed variation in survival expectations should be in accordance with epidemiological evidence regarding the relationships among mortality risk, health status, smoking and socioeconomic status. Table 2 indicates that for each target age T, the average SSP becomes dramatically lower as the health level decreases and as age

Table 2 Subjective survival probability in different categories of self-assessed health, smoking status, income and education

		Overall sample $N = 20{,}285$	$T = 75$ $N = 14{,}325$	$T = 80$ $N = 3{,}377$	$T = 85$ $N = 2{,}734$	$T = 90$ $N = 2{,}087$	$T = 95$ $N = 1{,}244$
Self-assessed health	Excellent	76.6 (0.477)	78.6 (0.494)	73.9 (1.548)	67.1 (2.137)	61.5 (3.698)	60.3 (7.456)
	Very good	69.8 (0.358)	73.0 (0.381)	67.1 (1.043)	61.9 (1.461)	50.3 (1.965)	45.7 (2.753)
	Good	62.8 (0.298)	68.0 (0.339)	62.4 (0.730)	54.6 (0.897)	47.3 (1.223)	38.6 (1.732)
	Fair	51.4 (0.447)	58.4 (0.596)	53.4 (1.043)	47.9 (1.135)	38.0 (1.267)	29.7 (1.562)
	Poor	39.3 (0.933)	48.3 (1.437)	41.8 (2.358)	37.3 (2.143)	23.6 (1.909)	24.1 (2.672)
Smoking status	Never smoker	62.5 (0.277)	68.6 (0.317)	61.5 (0.667)	53.1 (0.778)	42.7 (0.958)	33.3 (1.267)
	Current smoker	63.0 (0.438)	65.5 (0.473)	56.3 (1.301)	51.3 (1.818)	43.8 (3.145)	40.5 (3.781)
	Former smoker	61.2 (0.378)	70.3 (0.416)	61.8 (0.957)	53.3 (1.142)	41.8 (1.362)	36.3 (1.966)
Income	1st Income quartile	58.3 (0.412)	65.0 (0.523)	59.3 (0.957)	53.3 (1.063)	43.7 (1.296)	38.7 (1.804)
	2nd Income quartile	57.9 (0.417)	65.2 (0.531)	59.9 (0.922)	51.9 (1.061)	41.4 (1.335)	34.1 (1.712)
	3rd Income quartile	64.3 (0.381)	70.1 (0.407)	63.0 (0.994)	52.8 (1.260)	41.2 (1.644)	31.0 (2.371)
	4th Income quartile	67.4 (0.366)	70.9 (0.372)	62.0 (1.237)	55.2 (1.777)	43.8 (2.256)	31.3 (2.802)
Education	No education	56.0 (0.985)	63.6 (1.584)	57.1 (2.198)	55.1 (1.953)	41.7 (2.664)	43.8 (4.080)
	Less than diploma	59.1 (0.309)	66.7 (0.375)	60.1 (0.729)	52.4 (0.831)	42.5 (1.017)	34.2 (1.329)
	High school diploma	63.5 (0.347)	68.3 (0.380)	62.6 (0.909)	51.7 (1.243)	41.2 (1.585)	34.6 (2.129)
	University degree	66.9 (0.400)	71.2 (0.411)	61.2 (1.245)	56.2 (1.636)	45.0 (2.155)	33.2 (3.156)

The table reports average subjective survival probability (SSP) calculated in the overall sample and in sub-samples based on target age (T). Standard deviations in parentheses

increases (that is, as the target age moves from $T = 75$ to $T = 95$). Surprisingly, relative to never smokers, the average SSP at $T = 75$, 80 or 85 is higher for former smokers but lower for current smokers. The average SSP at $T = 90$ is lower for former smokers than for current and never smokers. In the overall sample, current and never smokers report higher average SSPs than former smokers. As expected, average SSP increases with income and education.

Additional analysis, reported in electronic supplementary material, shows that subjective health varies as expected with survival expectations; however, less clear-cut evidence appears on the link between smoking habit, survival expectations and subjective health (Table B.2). A marked relationship between parental death and the respondents' SSP, suggested by Hurd and McGarry (1995, 2002) emerges (Table B.3). Observed SSPs are also high for individuals with spouses who are still alive (or died during their fifties). These relationships provide motivation for including parental and spouse mortality as explanatory variables in our empirical model.

3 Model and estimation strategy

We propose a simultaneous recursive (triangular) model for survival expectations (SE_i), subjective health (H_i) and smoking behaviour (S_i); in this model, SE_i at any specific age depends on H_i and S_i, whereas H_i depends on S_i:

$$sE_i = f_E (H_i, S_i, X_i, \mu_E)$$
$$H_i = f_H (S_i, X_i, \mu_H)$$
$$S_i = f_S (X_i, \mu_S)$$

Our model, building on Grossman (1972) and Carbone et al. (2005), assumes that individuals assess their survival by weighing up both the direct effects of smoking on mortality risk (the long-term effects) and the indirect effects of smoking on their health (the short-term effects). We focus on two structural equations for reporting survival expectations and subjective health and a reduced-form equation for smoking duration. These three processes also depend on exogenous individual characteristics (X_i) and unobserved factors (μ_E, μ_H and μ_s, where the latter vector includes unobserved factors which influence survival expectations, subjective health and individual utility). To address unobservable heterogeneity in the econometric model, it is important to account for reverse causation and for endogeneity biases in the relationships among smoking, health perception and survival beliefs. As a result of reverse causation, the perception of particularly poor health and high mortality risks might decrease the probability and duration of smoking behaviours. The perception of smoking risks can be overestimated if this issue is not appropriately addressed. Subjective health and smoking behaviour can be endogenous to survival expectations, and smoking behaviour can be endogenous to subjective health. Endogeneity might arise from unobservable heterogeneity. In fact, unobserved factors (such as genetics, time preferences, risk aversion and the awareness of the health and mortality risks of smoking) might simultaneously affect the formation of survival expectations, reporting health and smoking behaviour.

Our approach to identification relies on the recursive triangular structure of the model, which imposes restrictions on parameters by construction, and non-linearity of functional forms of each equations. Exclusion restrictions, namely the omission of at least one variable in one equation, can be used to achieve more robust identification. We impose exclusion restrictions on each equation. In particular, from the two structural equations for survival expectations and subjective health we exclude an indicator of household composition measuring the number of children who live in the household.[12] This strategy is discussed in detail in the following pages where the smoking duration model is described.

We model survival expectations using a beta regression model[13]:

$$f(y_1 \mid y_2, t, s, q, x_1, \mu) = \frac{\Gamma(\omega + \tau)}{\Gamma(\omega)\Gamma(\tau)} y_1^{(\omega-1)}(1 - y_1)^{(\tau-1)} \tag{1}$$

where y_1 is (rescaled) SSP, t is smoking duration, s is a binary indicator that takes value 1 if the individual has ever started to smoke, q is a binary indicator that takes value of 1 if she has quit smoking and μ represents unobservable heterogeneity. Γ is the gamma distribution, and both ω and τ are shape parameters that define the precision parameter φ. Maximum Likelihood (ML) estimation is used (Paolino 2001). The expected value of SSP is approximated by a logistic model: $E(\text{SSP}) = \frac{\omega}{\omega+\tau} = \frac{\exp(z_1\beta)}{1+\exp(z_1\beta)}$, where z_1 includes subjective health, smoking and exogenous covariates (x_1). In particular, we use two dummy variables to distinctly identify current and former smokers and the interactions of these statuses with the number of years spent smoking.[14] Exogenous covariates include standard socioeconomic and demographic variables; indicators of lifestyle and objective health; and an indicator of optimism and country dummies. Lifestyle choices (drinking, physical exercise and obesity) can be regarded as health investment decisions that might mitigate smokers' risk perceptions and, therefore, affect perceived health and survival beliefs.[15] Our objective measures of health are hospital admissions and tobacco-related diseases in the previous twelve months; they are assumed to influence both SSP and SAH. The SHARE provides valuable informa-

---

[12] This indicator of household composition suffer from a remarkable problem of item non-response in wave 2, thus providing another motive for using wave 1 only.

[13] The beta distribution is often employed for proportions or subjective beliefs about the probabilities of future events (Ferrari and Cribari-Neto 2004; Smithson and Verkuilen 2006), because it well models continuous and bounded variables that are characterised by spikes at certain response foci. We rescale y_1 to lie in the interval $(0, 1)$. To avoid taking the logarithm of zeros and ones, $y_1 = \left(\frac{\text{SSP}}{100} \cdot (N-1) + a\right)$, where N is the sample size and a is a constant (in this case, $a = 0.5$). Alternative transformations can be utilised.

[14] The logarithm transformation of the continuous indicator is used to ensure flexibility in the relationship.

[15] Although lifestyle choices are likely endogenous to subjective health and survival expectations, our model does not control for this issue, but instead assumes the exogeneity of these factors. This assumption might represent a limitation of the analysis. As a robustness check, we have compared our model with another one that excludes lifestyles, but raises the issue of omitted variable bias. The coefficients of SAH and smoking in the SSP equation are very similar in magnitude for these two models; the signs and significance levels are unchanged. Information criteria favour the model that includes lifestyles. Results are available upon request.

tion about personality traits that allows us to use a binary indicator of pessimism.[16] Optimism appears to be central determinant of not only smoking patterns but also perception of the short- and long-term effects of smoking at different stages of the life cycle (more optimistic personalities typically tend to underestimate the risks of smoking). Other specific controls include indicators of parental and spouse mortality and an indicator of numeracy that captures cognitive ability. Moreover, to account for systematic differences in reporting expectations and the fact that respondents are not asked to evaluate their chances of survival for the same number of years, a continuous indicator of the difference between an individual's current age and target age is included.

Subjective health is modelled with a probit model that describes the probability of an individual reporting that she is in excellent, very good or good health:

$$\Pr(y_2 = 1 \mid t, s, q, x_2, \mu) = \Phi(z_2 \beta) \tag{2}$$

where y_2 is SAH and z_2 includes smoking and exogenous covariates (x_2).[17] The exogenous covariates are the same standard variables of Eq. (1). Disability indicators (i.e. gali, adl and iadl) are included because they might have a direct effect on perception of health quality and an indirect effect (through SAH) on survival beliefs, as individuals typically adapt quickly to the onset of disabilities and to sudden health changes.

Smoking behaviour is modelled using a two-part specification of the duration model, which implies splitting the sample according to the decision to start smoking (Douglas and Hariharan 1994; Forster and Jones 2001):

$$[\Pr(s = 1 \mid x_3, \mu) f(t \mid x_4, \mu)]^{s \cdot q} [\Pr(s = 1 \mid x_3, \mu) S(t \mid x_4, \mu)]^{s \cdot (1-q)}$$
$$[1 - \Pr(s = 1 \mid x_3, \mu)]^{(1-s)} \tag{3}$$

where a probit model for the probability to start smoking, $\Pr(s = 1 \mid x_3, \mu) = \Phi(x_3 \beta)$ describes the first part and a Weibull distribution, with density $f(t \mid x_4, \mu) = \lambda \alpha t_i^{(\alpha-1)} \exp(-\lambda t_i^{\alpha})$ and survival $S(t \mid x_4, \mu) = \exp(-\lambda t_i^{\alpha})$, describes the second part of the model, namely the hazard of quitting smoking.[18] Here α is the duration dependence parameter, whereas λ is $\exp(-x_4 \beta)$, a function of covariates.[19] In Eq. (3),

---

[16] This indicator of pessimism, which denotes whether a respondent reports that she has no hope for the future, is one of the items that is collected by the SHARE to define the EURO-D symptoms scale, which measures mental health.

[17] Dardanoni and Li Donni (2012) estimate a finite mixture model to explore the self-reporting effect of individual characteristics on SAH, and assume two latent classes of good health and bad health (in essence, SAH is conditional on an individual's unobservable true health). A similar framework is employed by our model, which assumes that true mortality risk and true health status of an individual directly go into the unobservable component of the model; thus, the latent classes are characterised by different states of unobservable health and true mortality risk.

[18] The Weibull has been chosen among other distributions on the basis of information criteria and Cox–Snell residuals test.

[19] The model is parameterised with respect to the accelerated failure time metric. The estimated coefficients should, therefore, be interpreted in terms of acceleration (or deceleration) of time until an individual ceases smoking.

x_3 includes only income, education and demographic variables, which are assumed to reflect prior socioeconomic characteristics that influence tobacco consumption; x_4 include standard exogenous variables and an indicator of household composition that measures the number of children who live in each respondent's households.[20] It is well known that the presence of children in a household might affect individuals' decisions about their smoking behaviours (see, e.g. Jarvis 1996); furthermore, it is reasonable to conjecture that the presence of children does not influence health perception and survival beliefs. Instruments for smoking behaviour should be correlated with smoking but uncorrelated with SAH and SSP. In the absence of information about past parental smoking habits, which could be a possible candidate for an instrument as discussed by Balia and Jones (2011), it appears difficult to find a good instrument for smoking behaviour over the life cycle (see Adda and Lechene 2012).

The sample likelihood of our recursive model which combines Eq. (1)–(3) is as follows:

$$
\begin{aligned}
L_i = {} & f\left(y_1 \,|\, y_2, t, s, q, x_1, \mu\right) \cdot \Pr\left(y_2 = 1 \,|\, t, s, q, x_2, \mu\right) \cdot \left[f\left(t \,|\, x_4, \mu\right)\right. \\
& \Pr\left(s = 1 \,|\, x_3, \mu\right)\right]^{s \cdot q} \cdot \left[S\left(t \,|\, x_4, \mu\right) \Pr\left(s = 1 \,|\, x_3, \mu\right)\right]^{s \cdot (1-q)} \cdot \\
& \left[1 - \Pr\left(s = 1 \,|\, x_3, \mu\right)\right]^{(1-s)}
\end{aligned}
\tag{4}
$$

In the presence of unobservable heterogeneity (μ) this likelihood is analytically intractable and an appropriate estimation approach is needed. We propose a finite mixture (FM) model that represents unobservable heterogeneity in terms of a finite number of latent classes, namely the types of individuals in the population from which the observed data are drawn (McLachlan and Peel 2000).[21] Given class membership, response variables are assumed to be independent of one another, so that a single response per individual is sufficient for identifying the model with cross-sectional data. Handling endogeneity in non-linear models can be challenging (Wooldridge 2002; Cameron and Trivedi 2005); FM models have been widely applied for this purpose because they assume the existence of correlated unobservable heterogeneity (see, e.g. Mroz 1999; Gilleskie and Strumpf 2005; Deb and Trivedi 2006).[22] A maximum likelihood (ML) estimation of our FM model is achieved through the use of the expectation-maximisation (EM) algorithm (see Appendix A in the electronic

---

[20] Smoking behaviour is strictly correlated with alcohol consumption, sedentary behaviour and dietary habits (Marcus et al. 1999; Di Novi 2010).

[21] FM models provide an alternative and flexible method of accounting for the unobserved individual effect in the case of non-linear models (see Cameron and Trivedi 2005). They have recently been used with panel data for pooled, population-averaged and random-effects models (see Bago d'Uva 2006; Bago d'Uva and Jones 2009) because of the following advantages: any parametric distribution for the unobservable heterogeneity is required and heterogeneity is allowed in the intercept as well as in the slopes. Deb and Trivedi (2013) have recently proposed a fixed-effect FM model for count data in short balanced panels (minimum $T = 4$) with exogenous regressors, using the EM algorithm. This type of extension to our model is problematic due to the availability of only two waves and the prospect of attrition bias.

[22] Mroz (1999) uses a latent factor model, which is a type of FM, to control for the endogeneity of a regressor. He shows that his methodology performs better than alternative estimators, particularly in the case of non-normality.

supplementary material). In principle, the non-linearity of the functional forms allows for each component of the mixture to be identified.

4 Results

4.1 Survival expectations

We discuss results obtained from the two-class model. Estimated beta regression coefficients and significance levels, reported in Table 3, are transformed in odds ratios, which are interpreted as the percentage change from the average SSP of the baseline individual.[23] Table 4 shows that the baseline individual expects to have a 82 % probability of survival at age 95 in the first latent class (class 1) and 44 % in the second latent class (class 2). Pessimism diminishes baseline survival probabilities, particularly in class 2 where pessimistic individuals exhibit average decreases in SSP of about 35 % (only 15 % among class 1 individuals). As expected, the average SSP decreases in both classes as the distance between the current and target age increases, and increases as the age class becomes younger. In class 1, the average SSPs are 139, 101 and 20 % higher than the baseline SSP for individuals who evaluate their survival at ages 75, 85 and 90; this effect is amplified in class 2 (the average SSPs are about 287, 184 and 33 % higher than the baseline SSP). We also find evidence of a statistically significant gender effect in reporting survival expectations: the baseline SSP decreases of about 8 % for males in both classes.

As expected, we find that SAH explains most of the observed variation in SSP. A better perceived health produces SSPs that are greater than in the baseline case by about 38 % for class 1 and 78 % for class 2. This large variation might be attributed to a greater diversity of true states of health for individuals in class 2. In class 1, the average SSP is significantly higher for former smokers (by 14.7 %) than for the baseline individual (a never smoker). The indicators of time spent smoking are statistically significant and negative in class 1, implying that SSP decreases with smoking duration but at a diminishing rate for each additional year spent smoking. This effect is larger for current (-6 %) than for former smokers (-4 %). For longer durations differences between current and former smokers tend to disappear: e.g. for durations of longer than 20 years the percentage change in SSP becomes very small (about -0.2) for both types of smokers, thus indicating the cognitive dissonance of the most addicted smokers.

Furthermore, the positive effect of quitting dominates the negative effect of smoking duration for former smokers. This might suggest that former smokers do not internalise the negative effects of their prior smoking on perceived mortality risk (an indication of a myopic attitude); instead, they reward themselves for quitting by assuming that they have better chances of living for a longer duration than their past smoking behaviours would appear to indicate. This result appears to suggest that the long-term effects of

---

[23] The baseline individual is a single female from the Netherlands who is between 81 and 85 years of age, is in fair or poor health, has never smoked, maintains a healthy lifestyle, earns a low level of income, has no education and is unemployed (or works as a housekeeper).

Table 3 Estimated coefficients from a Finite Mixture Model with two classes ($K = 2$) for subjective survival probability, subjective health and smoking

Equations	Survival expectations		Subjective health		Smoking duration		Starting smoking	
	Class 1	Class 2	Class 1	Class 2	Class 1	Class 2	Class 1	Class 2
Variables								
SAH	0.324***	0.578***						
Current smoker	0.103	−0.002	−0.430*	−0.393**				
Former smoker	0.137***	−0.092	0.254***	0.449***				
ln(sm. duration) × current smoker	−0.063**	−0.024	0.086	0.066				
ln(sm. duration) × former smoker	−0.041***	0.043*	−0.116***	−0.124***				
Drinker	0.018	−0.018	0.077***	0.062	0.161***	0.094*		
No physical exercise	−0.084***	−0.234***	−0.401***	−0.315***	0.254***	0.273***		
Obese	−0.034**	−0.011	−0.195***	−0.187***	−0.102***	−0.136***		
2nd Income quartile	−0.059**	−0.069	0.108*	−0.049	−0.077***	−0.056**	0.122***	0.064
3rd Income quartile	−0.013	0.059	0.166**	0.057	−0.153***	−0.057*	0.179***	0.067
4th Income quartile	0.019	0.011	0.193***	0.096**	−0.167***	−0.090*	0.105**	0.088
Less than diploma	−0.110***	0.050	0.112***	0.129***	0.175***	0.198***	0.156***	0.095
High school diploma	−0.058	0.102***	0.207***	0.281***	0.116**	0.123*	0.219***	0.207**
Univeristy degree	0.006	0.056	0.423***	0.403***	−0.007	−0.074	0.152	0.123
Retired	0.045***	0.049*	0.068**	0.117***	−0.098***	−0.006	0.834***	0.823***
Employed	0.095***	0.165***	0.361***	0.411***	−0.173***	−0.050**	0.383***	0.367***
Male	−0.086***	−0.079***	−0.025	−0.031	−0.004	0.051	0.177***	0.054
Age 50–65	0.873***	1.355***	−0.014	0.102	0.021	0.207***	0.083**	0.116
Age 66–70	0.700***	1.046***	0.037	0.081	0.043	0.114*	0.012	0.035
Age 71–75	0.454***	0.774***	−0.006	−0.054	−0.027	0.116*		

Table 3 continued

Equations	Survival expectations		Subjective health		Smoking duration		Starting smoking	
	Class 1	Class 2	Class 1	Class 2	Class 1	Class 2	Class 1	Class 2
Age 76–80	0.182***	0.285***	−0.060	−0.033	−0.135**	0.061		
Target age−age	−0.013***	−0.025***						
Married	0.052***	0.318***	0.041	0.126	−0.156***	−0.140*		
Separated or divorced	0.075**	0.141**	−0.044	0.036	0.032	0.152		
Widowed	−0.039	0.287***	0.026	0.115	−0.018	−0.002		
Partner age at death <50	0.069	0.097						
Partner age at death 51−74	0.005	−0.075						
Deceased mother	−0.118***	−0.139*						
Mother age at death 51−74	−0.026	0.046						
Mother age at death 75−79	0.027	0.020						
Mother age at death 80−84	0.061**	0.107						
Mother age at death 85−89	0.055	0.098						
Mother age at death >90	0.170***	0.265***						
Deceased father	−0.081***	−0.063						
Father age at death 51−74	−0.080***	−0.099**						
Father age at death 75−79	0.023	−0.061						
Father age at death 80−84	0.046***	0.004						
Father age at death 85−89	0.082***	−0.050						
Father age at death >90	0.117***	0.125**						
Numeracy	−0.001	0.016						
Pessimism	−0.165***	−0.436***	−0.282***	−0.455***	0.118***	0.090		

Table 3 continued

Equations	Survival expectations		Subjective health		Smoking duration		Starting smoking	
	Class 1	Class 2	Class 1	Class 2	Class 1	Class 2	Class 1	Class 2
Hospital admission	−0.085***	−0.165***	−0.291***	−0.425***				
Tobacco-related diseases	−0.092***	−0.120***	−0.454***	−0.409***	−0.127***	−0.029		
Adl			−0.289***	−0.348***				
Iadl			−0.159***	−0.047				
Gali			−1.169***	−1.238***				
Number of children				−0.020**	0.031***			
Intercept	0.009	−0.480***	1.412***	1.099***	4.002***	3.566***	−0.602***	−0.409***
ϕ	6.996***	0.456***		α	1.435***	1.457***		
$N = 20{,}285$	$\log L = -24005.935$		$p_1 = 0.66$					

Estimation method: Maximum Likelihood based on the EM algorithm, with cluster robust standard errors. Country dummies are included. Level of significance: *** 1 %; ** 5 %; * 10 %

Table 4 Estimated odds ratios for subjective survival probability and average partial effects for subjective health

	Survival expectations		Subjective health	
	Class 1	Class 2	Class 1	Class 2
SAH	1.383	1.782		
Current smoker	1.108	0.998	−0.100	−0.093
Former smoker	1.147	0.912	0.051	0.092
ln(sm. duration) × current smoker	0.939	0.977	0.019	0.015
ln(sm. duration) × former smoker	0.960	1.044	−0.025	−0.027
Drinker	1.018	0.982	0.016	0.014
No physical exercise	0.920	0.791	−0.095	−0.074
Obese	0.966	0.989	−0.043	−0.042
2nd Income quartile	0.943	0.933	0.024	−0.011
3rd Income quartile	0.987	1.061	0.037	0.013
4th Income quartile	1.019	1.011	0.042	0.021
Less than diploma	0.896	1.052	0.026	0.030
High school diploma	0.944	1.108	0.047	0.065
Univeristy degree	1.006	1.058	0.092	0.091
Retired	1.046	1.051	0.016	0.028
Employed	1.100	1.180	0.078	0.092
Male	0.918	0.924	−0.005	−0.007
Age 50−65	2.394	3.876	−0.003	0.023
Age 66−70	2.014	2.846	0.008	0.018
Age 71−75	1.575	2.168	−0.001	−0.012
Age 76−80	1.200	1.329	−0.013	−0.008
Target age−age	0.987	0.975		
Married	1.053	1.374	0.009	0.028
Separated or divorced	1.078	1.152	−0.010	0.008
Widowed	0.962	1.332	0.006	0.026
Partner age at death <50	1.072	1.101		
Partner age at death 51−74	1.005	0.927		
Mother died	0.888	0.870		
Mother age at death 51−74	0.974	1.047		
Mother age at death 75−79	1.028	1.020		
Mother age at death 80−84	1.062	1.113		
Mother age at death 85−89	1.056	1.103		
Mother age at death >90	1.186	1.303		
Father died	0.922	0.939		
Father age at death 51−74	0.923	0.905		
Father age at death 75−79	1.023	0.941		
Father age at death 80−84	1.047	1.004		
Father age at death 85−89	1.086	0.951		

Table 4 continued

	Survival expectations		Subjective health	
	Class 1	Class 2	Class 1	Class 2
Father age at death >90	1.124	1.133		
Numeracy	0.999	1.016		
Pessimism	0.848	0.646	−0.064	−0.108
Hospital admission	0.919	0.848	−0.067	−0.101
Tobacco-related diseases	0.912	0.887	−0.100	−0.093
Adl			−0.062	−0.077
Iadl			−0.034	−0.010
Gali			−0.306	−0.337
Number of children				
Baseline individual	0.821	0.442		

smoking are regarded as reversible. In class 2, only the indicator of smoking duration for former smokers has a statistically significant and positive coefficient. This finding can again be interpreted as a sign of the cognitive dissonance of smokers, who tend to think that lifestyles do not influence their mortality risk. The odds ratios measure a 4.4 % increase in the average SSP of former smokers relative to the baseline individual (a never smoker), but this effect diminishes as the duration of time spent smoking increases. Although the other smoking indicators are not significant, the sign of the coefficients suggests that the occurrence of any prior smoking behaviour decreases the chances of reporting a low SSP.

4.2 Subjective health

The impact of smoking and other individual characteristics on perceived health is measured by the average partial effects (APEs) of covariates on the probability of reporting a good, very good or excellent health status.[24] Table 4 indicates that pessimistic personalities tend to report poor perceived health in both latent classes. However, the effect of pessimism is stronger in class 2: pessimistic individuals are 11 % less likely than other individuals to report that their SAH is good (6.4 % in class 1). The probability of reporting good SAH increases monotonically with income and education and is higher for retired and employed individuals than for unemployed individuals; the APEs of income quartiles are larger in class 1, whereas the effects of education and occupational status are larger in class 2. Disability indicators have the expected negative effect and the APEs are larger in class 2, where, for example, the probability that SAH = 1 decreases by 33 % as the number of gali limitations increases. Individuals

---

[24] Partial effects are computed for each individual as the change in the probability that SAH = 1 caused by a change in a covariate, then averaged across the whole sample. Therefore, 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).

Table 5 Post-estimation predictions in hypothetical scenarios

	Class 1	Class 2
Average predicted SSP	58.1	66.1
In the smoking scenario	57.8	66.2
In the quitting scenario	57.9	64.9
In the smoking-free scenario	57.7	66.9
Average predicted probability of SAH $= 1$	0.73	0.71
In the smoking scenario	0.65	0.62
In the quitting scenario	0.79	0.79
In the smoking-free scenario	0.72	0.69
Average predicted probability of starting smoking	0.47	0.52
Average predicted smoking duration	38.80	39.25

Predicted average subjective survival probability (SSP) and average probability of good self-assessed health (SAH $= 1$) for hypothetical scenarios

who have been hospitalised during the prior year tend to report worse health than other individuals: this effect is larger in class 2 (11 %) than in class 1. For both classes, the presence of tobacco-related diseases is related to a 10 % lower probability of reporting good health. All lifestyle indicators are statistically significant in class 1; only alcohol consumption is associated with a higher probability of good SAH. In both classes, obesity and lack of physical exercise are associated with lower probabilities of good SAH. The effect of a sedentary lifestyle is larger in class 1 than in class 2.

The probability of reporting good health is significantly lower for current smokers (by about 10 % in both classes). By contrast, this probability is higher for former smokers (5 % in class 1 vs. 9 % in class 2). Time spent smoking is significant and negative only if interacted with past smoking behaviour; thus producing a reduction in the probability that SAH $= 1$ of about 3 % relative to never smokers. Non-linearity in the relationship between subjective health and smoking duration causes this effect to decrease as smoking duration increases. This evidence can again be interpreted as a signal that former smokers hold myopic attitudes and believe that the short-term effects of smoking are offset by the decision to quit (in other words, that the effects of smoking are reversible). The estimated APEs indicate that this finding is slightly stronger in class 2 than in class 1.

Table 5 shows that the average predicted SSP from the FM model is higher in class 2 than in class 1 (66 vs. 58 %). However, the average predicted probability of SAH $= 1$ is slightly higher in class 1 than in class 2 (73 vs. 71 %). We would have expected to find, instead, better SAH in the class of individuals with higher SSP. For each latent class, average predicted SSP and SAH are compared with predictions calculated for hypothetical scenarios involving various smoking behaviours. In the 'smoking-free scenario', nobody has ever smoked; in the 'quitting scenario', all individuals have quit smoking; and in the 'smoking scenario', all individuals are currently smoking. Independently of smoking, the average predicted SSP is always higher in class 2, with the highest value (66.9) in the 'smoking-free scenario'. In class 1, the highest average value of predicted SSP (57.9) is associated with the 'quitting scenario'. For

both classes, the highest average predicted probability of SAH $= 1$ is calculated for the 'quitting scenario'. In particular, this probability is about 13 and 17 % points greater in the 'quitting scenario' than in the 'smoking scenario', and about 6 and 10 % points larger than in the 'smoking-free scenario', for class 1 and class 2, respectively. For both classes, we predict that the lowest average predicted probability of SAH $= 1$ will occur in the 'smoking scenario'. These simulations appear to suggest that quitting smoking alters the risk perception of smokers, who tend to report better evaluations of survival expectations and health status than current smokers.

4.3 Smoking behaviour

Results in Table 3 show that the propensity to become a smoker is higher for men in both classes. In class 1, this propensity is higher for individuals from the youngest age cohort and is positively related to income and education, but coefficients do not show a clear-cut socioeconomic gradient. Estimates reflect heterogeneity across classes with respect to the hazard of quitting. Smoking duration is predicted to be significantly shorter for richer, retired and employed smokers, particularly in class 1. Educated individuals, notably in class 2, smoke for longer durations but this positive effect is predicted to decrease as education levels increase. In both classes, smokers with an unhealthy lifestyle in terms of drinking and exercise tend to quit later, whereas smokers with bad dietary habits quit more quickly. The effects of obesity and physical exercise are higher in class 2. Class 1 smokers who report that they have experienced tobacco-related diseases tend to quit earlier, but this effect is not observed in class 2. Marriage appears to protect smokers from a long history of tobacco use and this effect is more prominent in class 1. The indicator of pessimism is statistically significant and positive only in class 1, suggesting that pessimistic smokers, who face a low opportunity cost of smoking, are predicted to quit later in life. As reported in Table 5, the predicted probability of starting to smoke is higher in class 2 (0.52) than in class 1 (0.47), and the estimated average smoking duration is also longer in class 2 (39.3 years) than in class 1 (38 years).

The FM model is first estimated with two latent classes ($K = 2$); subsequently, the number of classes is augmented, and the statistical fit of the alternative models is compared through the use of information criteria. The FM model are also compared with the standard ML model that relaxes the assumption of unobservable heterogeneity. Table 6 shows that the FM model with $K = 2$ has a better fit than the ML model. Only the Akaike information criterion (AIC) suggests that the FM model performs better with $K = 3$. The consistent AIC (CAIC) and the Bayesian information criterion (BIC), which penalises the number of estimated parameters more severely than the AIC, are lower in the model with $K = 2$ thus supporting our choice.

4.4 Posterior analysis

The estimation results reveal that there are two unobservable populations that differ in the hazard of quitting and in the ways in which they formulate survival expectations and report subjective health. We can use the estimated class membership probability to describe the types of individuals who might belong to each class. We utilise a cut-

Table 6 Comparison of standard Maximum Likelihood model and Finite Mixture models using information criteria

	Standard ML model $K = 1$	FM model $K = 2$	FM model $K = 3$
AIC	61473.91	48306.87	47235.54
CAIC	62638.80	50938.57	51187.05
BIC	62637.80	50937.57	51186.05
No. parameters	147	295	443
N	20285	20285	20285

K is the number of classes or components of the mixture. The consistent Akaike information criterion (CAIC) is calculated as $2 \log L + [1 + \log(N)q]$ where q is the number of parameters

off probability of 0.5 to assign each individual to the class associated with a larger posterior probability and then define a binary indicator of class membership (d_i) that takes a value of 1 if the posterior probability is above the cut-off probability. Because $\sum_{k=1}^{K} \pi_{ik} = 1$, this approach is equivalent to stating that individual i belongs to latent class 1 if the π_{i1} is larger than π_{i2}. We estimate a probit model for the class membership probability conditional on the examined covariates and outcomes.[25] Table 7 shows that individuals who report higher SSPs and current or former smokers tend to have a lower probability of belonging to class 1 (which generally represents less addicted smokers than class 2). Current smokers have a 13 % lower probability of being in class 1 than never or former smokers. The following characteristics also appear to be associated with significantly lower probability of membership in class 1: pessimistic personalities; retirees; married or separated marital status; a lack of regular exercise; the occurrence of at least one hospital admission in the last year; and a deceased father. Individuals who report better SAH but experience limitations in their usual activities due to health problems (gali) are more likely to belong to class 1. Class membership significantly increases with age. This effect is smaller for individuals in age classes 71–75 and 76–80, thus implying that the oldest individuals are more likely to belong to the class of more addicted smokers. This finding, which is in line with Adda and Lechene (2012), could be interpreted as representative of 'rationally myopic' behaviour. For elderly individuals, who have relatively little remaining lifespan, the opportunity cost of smoking is low due to the minimal effects on mortality risk.

5 Conclusion

This work aims to assess how smokers' risk perceptions affect survival expectations and subjective health. The analysis investigates perceptions of the short- and long-term effects of smoking and examines whether smokers believe that the detrimental effects of smoking can be reversed. The empirical literature that addresses this topic, primarily based on US surveys, provides mixed evidence regarding the relationship

---

[25] Table B.4, available in the electronic supplementary material, compares individuals in sub-sample 1, drawn from the first population, with individuals in sub-sample 2. We find that these two groups of individuals differ in terms of observable characteristics and particularly in terms of smoking behaviour. Such differences are in line with the FM model results that we have previously discussed.

Table 7 Class membership probability

Variables	Coefficients	Average partial effects
SSP	−0.011***	−0.003
SAH	0.221***	0.070
Current smoker	−0.409*	−0.131
Former smoker	−0.137*	−0.041
ln(sm. duration) × current smoker	0.062	0.019
ln(sm. duration) × former smoker	0.006	0.002
Drinker	−0.036	−0.011
No physical exercise	−0.223***	−0.073
Obese	−0.051	−0.016
2nd Income quartile	−0.006	−0.002
3rd Income quartile	0.074*	0.023
4th Income quartile	0.049	0.015
Less than Diploma	−0.049	−0.015
High school diploma	−0.033	−0.010
Univeristy degree	0.016	0.005
Retired	−0.047*	−0.014
Employed	0.010	0.003
Male	−0.020	−0.006
Age 50−65	0.536***	0.182
Age 66−70	0.572***	0.193
Age 71−75	0.436***	0.152
Age 76−80	0.191***	0.069
Target age−age	−0.001	0.000
Married	−0.112**	−0.034
Separated or divorced	−0.123***	−0.037
Widowed	−0.032	−0.009
Partner age at death <50	0.081	0.024
Partner age at death 51−74	−0.074	−0.023
Deceased mother	−0.068	−0.021
Mother age at death 51−74	0.049	0.015
Mother age at death 75−79	0.071	0.022
Mother age at death 80−84	0.062	0.019
Mother age at death 85−89	0.032	0.010
Mother age at death >90	0.051	0.016
Deceased father	−0.062*	−0.019
Father age at death 51−74	0.033	0.010
Father age at death 75−79	0.035	0.011
Father age at death 80−84	0.058	0.018
Father age at death 85−89	0.027	0.009
Father age at death >90	0.056	0.017

Table 7 continued

Variables	Coefficients	Average partial effects
Numeracy	0.046***	0.014
Pessimism	−0.124***	−0.039
Hospital admission	−0.085*	−0.027
Tobacco-related diseases	−0.006	−0.002
Adl	−0.031	−0.009
Iadl	−0.034	−0.011
Gali	0.043**	0.013
Number of children	−0.003	−0.001
Intercept	0.921***	
$N = 20,285$	$\log L = -10979.458$	

Estimated coefficients and average partial effects.
Dependent variable: binary indicator of class membership d_i (takes a value of 1 if respondents belong to the sub-sample drawn from the population of the first type). Estimation method: probit model, Maximum Likelihood with cluster robust standard errors. Country dummies are included. Level of significance: *** 1 %; ** 5 %; * 10 %

between smoking and risk perception. In this study, we use data on elderly Europeans from the SHARE to explore the formation of individuals' self-assessments of their own mortality risk and health. We estimate a joint recursive model for survival expectation, subjective health and smoking duration within a finite mixture approach, which addresses the potential endogeneity of the health and smoking variables and the reverse causation that arise from unobservable heterogeneity. We provide evidence of heterogeneity in assessing subjective survival probability and reporting subjective health, and show that two unobservable types of individuals can be identified in the examined population. After controlling for observed characteristics, including optimism, it appears that the remaining differences in both the smoking patterns and perceptions of smoking effects between the two types of individuals may reasonably depend on unobservable factors, such as information constraints, which may be regarded as limited awareness of the consequences of smoking; risk aversion; time preferences; life experiences; genetics; true health and true mortality risk (unobservable frailty).

We find that the population of the first type is characterised by lower average survival expectation and higher average probability of reporting good health and includes a smaller proportion of smokers than the population of the second type. Because of longer smoking histories, we can refer to the latter population as the individuals who are more addicted to smoking. For the population of the first type, smoking is an important predictor of survival expectations, and smokers incorporate the (long-term) risks of smoking into their assessments of survival probabilities. However, for both current and former smokers, this effect vanishes as smoking duration increases (i.e. for the most addicted individuals). Furthermore, we find that among individuals of the first type, former smokers are present-oriented (myopic) and believe that smoking effects are reversible. In fact, the negative effect of smoking duration on survival beliefs is counterbalanced by a decision to quit smoking; as a result, former smokers largely overestimate their survival probabilities. By contrast, individuals of the second

type do not incorporate the negative effects of smoking duration into their survival beliefs; among these individuals, former smokers significantly overestimate their survival expectations, although this occurs to a diminishing degree as smoking duration increases. In both unobservable types of individuals, subjective health is significantly lower for current smokers. Time spent smoking produces a reduction in subjective health only for former smokers, but this effect goes to zero as duration increases. Furthermore, for former smokers, the short-term effects of smoking duration are subjectively offset by the decision to quit smoking. This evidence is common to both classes and is stronger for individuals in the second population type. The observed attitudes regarding risk perception might relate to the fact that smoking involves low opportunity costs for individuals with lower life expectancies; this consideration may be particularly salient for the oldest individuals, who are more concentrated in the second population.

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 the morbidity and mortality consequences of smoking. This paper demonstrates that despite the existence of various national anti-smoking campaigns in Europe, smokers are not necessarily fully aware of the true health risks that smoking produces. This raises the question of whether information about smoking should be more widely disseminated or whether, using a concept from behavioural economics, it should be more salient to improve its impact on reducing smoking.

Furthermore, the perception of the reversibility of the smoking effects reflects the myopic behaviours of former smokers. This issue presents a second question of whether the dissemination of more detailed information is required. Results suggest that the targeting of heterogeneous types of individuals can be an important element of smoking-related policy interventions. Our analysis also reveals that the numerical indicator of subjective survival probability is a complete measure of survival expectations that captures both observable and unobservable factors that influence individual beliefs. The use of this indicator in models of health investments over the life-cycle might be preferred to the use of mortality hazard rates estimated from life tables; this prospect provides ample scope for future research work.

References

Adda J, Lechene V (2012) Health selection and the effects of smoking on mortality. IZA discussion papers 6206

Arcidiacono P, Sieg H, Sloan F (2007) Living rationally under the volcano? An empirical analysis of heavy drinking and smoking. Int Econ Rev 48(1):37–65

Bago d'Uva T (2006) Latent class models for use of health care. Health Econ 15:329–343

Bago d'Uva T, Jones AM (2009) Health care utilisation in Europe: new evidence from the ECHP. J Health Econ 28:265–279

Balia S, Jones AM (2011) Catching the habit: a study of inequality of opportunity in smoking-related mortality. J R Stat Soc A 174(1):175–194

Becker GS, Murphy K (1988) A theory of rational addiction. J Polit Econ 96(4):675–700

Becker GS, Grossman M, Murphy KM (1994) An empirical analysis of cigarette addiction. Am Econ Rev 84(3):396–418

Bruine de Bruin W, Fischbeck PS, Stiber NA, Fischhoff B (2002) What number is "fifty–fifty"?: redistributing excessive 50 % responses in elicited probabilities. Risk Anal 922(4):713–723

Cameron AC, Trivedi PK (2005) Microeconometrics. Methods and applications. Cambridge University Press, New York

Carbone JC, Kverndokkb S, Røgebergb OJ (2005) Smoking, health, risk, and perception. J Health Econ 24:631–653

Cawley J, Ruhm CJ (2011) Chapter Three-the economics of risky health behaviors. In: Mark V Pauly, Thomas G Mcguire, Pedro P Barros (eds) Handbook of Health Economics, vol 2. Elsevier, Amsterdam, pp 95–199

Chaloupka F (1991) Rational addictive behavior and cigarette smoking. J Polit Econ 99(4):722–742

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

Cutler DM, Glaeser EL (2009) Why do Europeans smoke more than Americans? In: Wise DA (ed) Developments in the economics of aging. NBER book, Cambridge, MA, pp 255–282

Dardanoni V, Li Donni P (2012) Reporting heterogeneity in health: an extended latent class approach. Appl Econ Lett 19(12):1129–1133

Deaton AS, Paxson C (1998) Ageing and inequality in income and health. Am Econ Rev Papers Proc 88:248–253

Deb P, Trivedi PK (2006) Specification and simulated likelihood estimation of a non-normal treatment-outcome model with selection: application to health care utilization. Econom J 9:307–331

Deb P, Trivedi PK (2013) Finite mixture for panels with fixed effects. J Econom Methods 2(1):35–51. doi:10.1515/jem-2012-0018

Delavande A, Rohwedder S (2011) Differential survival in Europe and the United States: estimates based on subjective probabilities of survival. Demography 48:1377–1400

Dempster AP, Laird N, Rubin DB (1977) Maximum likelihood from incomplete data via EM algorithm. J R Stat Soc B 34(1):1–38

Di Novi C (2010) The influence of traffic-related pollution on individuals' life-style: results from the BRFSS. Health Econ 19:1318–1344

Dominitz J, Manski CF (1997) Using expectation data to study subjective income expectations. J Am Stat Assoc 92:855–867

Douglas S, Hariharan G (1994) The hazard of starting smoking: estimates from a split population duration model. J Health Econ 13(2):213–230

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

Ferrari S, Cribari-Neto F (2004) Beta regression for modelling rates and proportions. J Appl Stat 31(7):799–815

Forster M, Jones AM (2001) The role of tobacco taxes in starting and quitting smoking: duration analysis of British data. J R Stat Soc A 164:517–547

Gilleskie D, Strumpf KL (2005) The behavioral dynamics of youth smoking. J Hum Resour 40:822–866

Grossman M (1972) On the concept of health capital and the demand for health. J Polit Econ 80(2):223–255

Guiso L, Tiseno A, Winter J (2005) Expectations. In: Börsch-Supan A, Brugiavini A, Jürges H, Mackenbach J, Siegrist J, Weber G (eds) Health, Ageing, and Retirement in Europe. Mannheim, Mannheim Research Institute for the Economics of Aging, pp 332–338

Hammar H, Johansson-Stenman O (2004) The value of risk-free cigarettes-do smokers underestimate the risk? Health Econ 13:59–71

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

Hurd M, McFadden D, Merrill A (1999) Predictors of mortality among the elderly. NBER working paper 7440

Hurd M, McGarry K (1995) Evaluation of the subjective probabilities of survival in the health and retirement study: data quality and early results. J Hum Resour 30:268–S292

Hurd M, McGarry K (2002) The predictive validity of subjective probabilities of survival. Econ J 112:966–985

Jacho-Chávez DT, Trivedi PK (2009) Computational considerations in empirical microeconometrics: selected examples. In: Mills TC, Patterson K (eds) Palgrave handbook of econometrics, vol 2. Applied Econometrics. Part V. Palgrave Macmillan, Basingstoke

Jarvis MJ (1996) The association between having children, family size and smoking cessation in adults. Addiction 91(3):427–434

Jürges H (2007) True health vs response styles: exploring cross-country differences in self-reported health. Health Econ 16:163–178

Khwaja A, Sloan F, Chung S (2007) The relationship between expectations and behaviours: mortality expectations and smoking decisions. J Risk Uncertain 35:179–201

Manski CF (2004) Measuring expectations. Econometrica 72(5):1329–1376

Marcus BH, Lewis BA, Hogan J, King TK, Albrecht AE, Bock B, Parisi AF, Niaura R, Abrams DB (1999) The efficacy of exercise as an aid for smoking cessation in women: a randomized controlled trial. Arch Intern Med 159:1229–1234

McLachlan GJ, Krishnan T (1996) The EM algorithm and extensions. Wiley, New York

McLachlan J, Peel D (2000) Finite mixture models. Wiley, New York

Michaud PC, Goldman D, Lakdawalla D, Gailey A, Zheng Y (2011) Differences in health between Americans and Western Europeans: effects on longevity and public finance. Soc Sci Med 73:254–263

Mroz TA (1999) Discrete factor approximations in simultaneous equation models: estimating the impact of a dummy endogenous variable on a continuous outcome. J Econom 92:233–274

Paolino P (2001) Maximum likelihood estimation of models with beta-distributed dependent variables. Polit Anal 9(4):325–346

Peracchi F, Perotti V (2010) Subjective survival probabilities and life tables: evidence from Europe. EIEF working paper 10/16

Schafer JL (1997) Analysis of incomplete multivariate data, vol 72. In: Monographs on statistics ad applied probability. Chapman and Hall, London

Schoenbaum M (1997) Do smokers understand the mortality effects of smoking? Evidence from the health and retirement survey. Am J Public Health 87(5):755–759

Smith VK, Taylor DH, Sloan FA (2001a) Longevity expectations and death: can people predict their own demise? Am Econ Rev 91:1126–1134

Smith VK, Taylor DH, Sloan FA, Johnson FR, Desvousge WH (2001b) Do smokers respond to health shocks? Rev Econ Stat 83(4):675–687

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

Thaler R, Shefrin H (1981) An economic theory of self-control. J Political Econ 89(2):392–406

van Doorslaer E, Gerdtham UG (2003) Does inequality in self-assessed health predict inequality in survival by income? Evidence from Swedish data. Soc Sci Med 57:1621–1629

Viscusi WK (1990) Do smokers underestimate risks? J Polit Econ 98(6):1253–1269

Viscusi WK (1991) Age variations in risk perceptions and smoking decisions. Rev Econ Stat 73(4):577–588

Viscusi WK, Hakes JK (2008) Risk beliefs and smoking behavior. Econ Inq 46(1):45–59

WHO (2011) Noncommunicable diseases country profiles 2011. http://www.who.int

Winston G (1980) Addiction and backsliding: the theory of compulsive consumption. J Econ Behav Organ 1:295–324

Winter J (2008) Expectations and attitudes. In First results from the Survey of Health, Ageing and Retirement in Europe (2004–2007). In: Börsch-Supan A, Brugiavini A, Jürges H, Kapteyn A, Mackenbach J, Siegrist J, Weber G (eds) Starting the longitudinal Dimension. MEA, Mannheim, pp 306–311

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