IIASA
International Institute for
Applied Systems Analysis
www.iiasa.ac.at

# Smoking, Education and the Ability to Predict Own Survival Probabilities: An Observational Study on US Data

**Arpino, B., Bordone, V. & Scherbov, S.**

**IIASA Working Paper**

**WP-17-012**

**Approved by: Wolfgang Lutz, Program Director World Population Program**
**30 June 2017**

International Institute for
Applied Systems Analysis
Schlossplatz 1
A-2361 Laxenburg, Austria

Tel: +43 2236 807 342
Fax: +43 2236 71313
E-mail: repository@iiasa.ac.at
Web: www.iiasa.ac.at

**Working Paper** **WP-17-012**

# Smoking, Education and the Ability to Predict Own Survival Probabilities: An Observational Study on US Data

Bruno Arpino (bruno.arpino@upf.edu)
Valeria Bordone (valeria.bordone@lmu.de)
Sergei Scherbov (scherbov@iiasa.ac.at)

**Approved by**

Wolfgang Lutz
Program Director, World Population Program
June 2017

# Contents

## Abstract

Background: Subjective survival probabilities (SSPs) are a good predictor of mortality, go beyond the aggregate description of survival defined by life tables, and are important for individuals' decision-making in later life. Despite the well-known mortality differentials by education as well as by characteristics such as smoking, little investigation has focused on SSPs by population sub-groups.

Methods: We use data on individuals aged 50-89 from the Health and Retirement Study (HRS) carried out in the USA between 2000 and 2012 (N=23,895). Each respondent was asked to assess the probability to survive to a given target age according to their age at the time of the survey. We assess how individuals' SSPs and estimated objective survival probabilities (OSPs) vary by education and smoking and calculate, for each respondent, the gap between them.

Results: Consistently with real mortality patterns, smokers report the lowest SSPs, both among lower and higher educated people. When comparing SSPs and OSPs we find that, irrespectively of the smoking status, higher educated people are more likely to correctly predict their survival probabilities than their lower educated counterparts. Within both education groups, past smokers better predict their survival probability. Current smokers with low education show the highest probability to overestimate their survival probability.

Conclusions: Lower educated people and smokers are aware of their lower life expectancy. Still, they overestimate their survival probabilities more than the higher educated and non-smokers. Our findings emphasize the need for policy makers to disseminate information about the risks of smoking, targeting people with lower education.

# Acknowledgments

The work was supported by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC under Grant ERC2012-AdG 323947-Re-Ageing.

## About the Authors

Bruno Arpino is Associate Professor at the Department of Political and Social Sciences and co-director of The Research and Expertise Centre for Survey Methodology (RECSM), Universitat Pompeu Fabra.

Valeria Bordone is Lecturer at the Department of Sociology, University of Munich (LMU). She is also Guest Research Scholar with the World Population Program at IIASA, Wittgenstein Centre for Demography and Global Human Capital (IIASA, VID/ÖAW, WU).

Sergei Scherbov is Deputy Program Director and Project Leader of the Reassessing Aging from a Population Perspective Project (Re-Aging) at the World Population Program at IIASA, and Leader of the Population Dynamics and Forecasting Group at the Vienna Institute of Demography (VID) of the Austrian Academy of Sciences, Wittgenstein Centre for Demography and Global Human Capital (IIASA, VID/ÖAW, WU). He is also Head of International Laboratory on Demography and Human Capital at the Russian Presidential Academy of National Economy and Public Administration (RANEPA).

# Smoking, Education and the Ability to Predict Own Survival Probabilities: An Observational Study on US Data

Bruno Arpino
Valeria Bordone
Sergei Scherbov

## 1 Introduction

Subjective survival probability (SSP) survey question – i.e., the probability a person assigns to the likelihood to survive to a certain age – is a good predictor of mortality. Although having precise expectations on the own survival curve is hard to achieve, previous studies show that SSPs predict mortality well even after controlling for mortality-related risk factors, such as objective health measures (Manski 2004; Perozek 2008; Hurd & McGarry 1995; Hurd & McGarry 2002; Siegel et al. 2003; van Doorn & Kasl 1998; Smith et al. 2001; Elder 2013; Kutlu-Koc & Kalwij 2017). SSPs contain subtle information, incorporating people's knowledge about their characteristics and behaviours that is not fully captured by standard health and risk behaviour measures. Methodologically, SSP questions have been proved to hold good properties, i.e., they are directly comparable across individuals (Dominitz & Manski 1997).

The literature on SSPs using data on the USA from the Health and Retirement Study (HRS) shows that they are, in general, consistent with the observed survival patterns at the population- and individual-level (Hurd & McGarry 2002; Siegel et al. 2003; Smith et al. 2001; Hurd 2009; Novak & Palloni 2013). Yet, sub-groups within the population may not only display different survival probabilities depending on both observed and unobserved characteristics, but also be more or less able to predict the own survival.

Among the studies assessing how SSPs vary across individuals (Perozek 2008; Khwaja et al. 2007; Ludwig & Zimper 2013; Bissonnette et al. 2014) several have shown that they vary with observed individual characteristics such as health, parental longevity, BMI, and smoking (Hurd & McGarry 1995; Hurd & McGarry 2002; Kutlu-Koc & Kalwij 2017; Falba & Busch 2005). In view of the high percentage of the American population that consists of current or past smokers, a percentage that reached 77% in some male cohorts (Wang & Preston 2009), this study focuses on the smoking behaviour.

Given that smoking increases the risk of numerous diseases (Taghizadeh et al. 2016) and that people are usually aware of a higher mortality risk associated with smoking (Balia 2014; Wang 2014), we expect smoking to be negatively correlated with SSPs.

While it is clear that education affects mortality (Cacciani et al. 2015; Bijwaard et al. 2015), most of the literature assessing SSPs has ignored the role of education in the ability to estimate survival probabilities and educational differences in the smoking behaviour (Gilman et al. 2008). Therefore, we analyse the ability of estimating own survival by smoking behaviour and education simultaneously.

This study first aims at identifying differences in the subjective perceptions of survival among lower and higher educated, according to whether they currently smoke, they have smoked in the past, or they have never smoked. To do so, we compare SSPs obtained from a population survey by education and smoking status, using data from HRS. Second, we check whether the results obtained from the analyses of the subjective survival probabilities reflect the real data, i.e. the actual survival patterns of the sub-populations considered. Because this information is not available in life tables, which at best contain average survival probabilities for a few population sub-groups, we estimate objective survival probabilities (OSPs) from the longitudinal sample of HRS using a Gompertz survival model. Finally, we compare subjective and objective survival probabilities across sub-groups, defined by education and smoking behaviour.

Understanding the variability of SSPs within a population is of high relevance because they affect life-cycle decisions under uncertainty (Hurd 2009). For example, individuals are likely to take important decisions, such as when to exit the labour market or whether to buy a life insurance, also based on their longevity expectations. SSPs have also been found to influence subsequent healthy behaviours (Novak & Palloni 2013). Moreover, if smokers correctly assess the risk of smoking and its influence on survival, their choice of continuing smoking will be a conscious one. Yet, in case of a considerable underestimation of the risk of smoking, overestimating the expected survival, individuals and their families may face negative consequences in terms of, for example, unexpected reduced household income because of illness or death. We acknowledge that overestimating the own survival may also bring together some (short-term) positive psychological effects. However, an analysis of the consequences of incorrect survival expectations goes beyond the scope of this paper, where we focus on whether different groups of people systematically differ in their ability to predict survival.

# 2 Methods

## 2.1 Data

We use data from the Health and Retirement Study (HRS), an age-cohort-based longitudinal panel survey of persons aged 50 and older in the USA. We consider respondents interviewed for the first time in 2000, 2002, 2004, 2006, 2008, 2010, and 2012 waves.

Our analysis applies to older adults aged 50–89 years old. It excludes respondents living in nursing homes as their number is limited to carry out separate analyses, yet they might substantially differ from the others in terms of their (health) conditions. We also exclude respondents interviewed in 2013 because their exposure time was too short. Moreover, most of the interviews of the last wave were held during the year 2012. The remaining working sample included 23,895 respondents.

## 2.2. Outcome variables

*Subjective survival probability (SSP)*: Since the 2000 wave, SSP has been asked among self-respondents aged 50–89 years old as follows: “I would like for you to give me a number from 0 to 100, where 0 means that you think there is absolutely no chance, and 100 means that you think the event is absolutely sure to happen. What is the percent chance that you will live to be [75 (if age is 65 or less) /80 (if age is 66–69) /85 (if age is 70–74) /90 (if age is 75–79) /95 (if age is 80–84) /100 (if age is 85–89)] or more?" (for details see http://hrsonline.isr.umich.edu/index.php?p=qnaires). For respondents aged 50–65, SSP was first asked with the target age of 75; unless they reported 0, it was then asked with 80 years as the target age. We used only the first target in the analysis to have one target age per person.

*Objective survival probability (OSP)*: Benefiting from the longitudinal nature of HRS, we know whether a person died after first interview until the year 2013. In addition to the information on vital status obtained through tracking of respondents, the HRS matches to the National Death Index for persons who are reported as deceased or who are not known to be alive through contact. For all submitted cases that were flagged as valid by the National Center for Health Statistics (NCHS) and verified by HRS staff, the Tracker file contains year and month of death, match score, and an alive/deceased flag.

### 2.3 Explanatory variables

*Education* distinguishes between higher (at least master degree) and lower attainment. *Smoking* behaviour considers whether the respondent has never smoked; has been a smoker in the past (at least 100 cigarettes – as specified in the HRS question); or currently smokes cigarettes.

### 2.4 Control variables

We control for *age*, *gender*, *ethnicity* (White/Caucasian; Black/African American; other) and *health* (i.e., whether the respondent was ever diagnosed with cancer, stroke, lung problems, and/or heart disease). In the model predicting SSPs we also control for dummy variables for the *target age* and interactions between these and respondent's age.

## 3 Statistical analysis

In a first step, we used linear regression models to analyse the association of smoking behaviours and education with SSPs. As the outcome variable (SSP) is bounded at 0 and 100, we also carried out the analyses using a Tobit regression model. Although these gave very similar results, we preferred the linear approach for simplicity of interpretation. From this model, we obtained adjusted survival probabilities for different sub-groups (i.e., considering smoking behaviour and educational attainment) and plotted them to ease interpretation of results. The regression model we estimated has the following form:

$$SSP = \beta_0 + \beta_1 Past + \beta_2 Current + \beta_3 High\_edu + \beta_4 Past * High\_edu + \beta_5 Current * High\_edu + \alpha X + \varepsilon \quad (1)$$

where *Past*, *Current* and *High_edu* are our main independent variables representing, respectively, dummy variables for past smokers, current smokers, and higher educated. People who never smoked and those with lower education represent the reference categories, respectively. Interactions between smoking and education are also included.

$X$ represents the control variables listed above and $\varepsilon$ is the error term. It should be noted that SSPs reported by the respondents can be interpreted as probabilities to survive to the target age, conditional to survival to the age observed at the time of interview (Hurd & McGarry 2002).

In a second step, we applied a survival model to real mortality data (i.e., survey data on whether each individual survived to 2013) to assess the association between our explanatory variables and objective mortality. We used a Gompertz model, that is widely used to model human mortality and it has been found to fit survival data of humans aged 10 to at least 85 better than the alternative survival functions (Wilson 1994; Kutlu-Koc & Kalwij 2017). Let $T$ indicate a random variable representing the respondent's age at death (life duration). Each respondent $i$ is aged $t_{0,i}$ at the start of the observation period (first interview) and aged $t_i$ at the end of the observation period (2013) or at the time of death, whichever comes first. In a Gompertz model the survivor function takes the following form:

$$S(t) = \exp\left\{-\lambda\gamma^{-1}\left(e^{\gamma t} - 1\right)\right\}.$$

The model is implemented by parameterizing $\lambda_i = exp(Z_i\beta)$, $Z$ being the vector of independent variables (including the interactions between smoking behaviour and education), implying that $h_0(t) = exp(\gamma t)$, where $\gamma$ is an ancillary parameter to be estimated from the data. This model provides the estimates of OSPs by smoking and education. In a third step, we compare, for each respondent, SSPs and the corresponding predicted OSPs conditional to the baseline age. To do so, we calculate the logarithm of the absolute difference between OPS and SSP and regress it on smoking, education, and the control variables. Then, we create a categorical variable Y as follows:

$$Y_i = \begin{cases} 1\,if & \left|\dfrac{SSP_i - OSP_i}{OSP_i}\right| * 100 \leq 10; \\ 2\,if & \left|\dfrac{SSP_i - OSP_i}{OSP_i}\right| * 100 > 10;\ SSP_i > OSP_i; \\ 3\,if & \left|\dfrac{SSP_i - OSP_i}{OSP_i}\right| * 100 > 10;\ SSP_i < OSP_i. \end{cases}$$

which indicates whether the respondent could reasonably approximate the own real OSP ($Y = 1$), "overestimated" it ($Y = 2$), or "underestimated" it ($Y = 3$). A reasonable estimate ("correct") is defined as an absolute gap between SSP and OSP not bigger than 10%. In a final step we used a multinomial logistic regression model with $Y$ as dependent variable and *Past*, *Current*, *High_edu,* their interactions, and $X$ as independent variables to estimate the probability to correctly estimate / overestimate / underestimate OSPs by smoking and education.

Differently from previous studies that compared SSPs to real mortality data using life tables (e.g. (3, 4, 19)), we compared SSPs to OSPs estimated on the same survey data for two reasons. First, life tables are only available at the population-level or for a few sub-groups of the population. Second, SSPs are estimated on the same selected sample (HRS respondents who survived up to the interview date, who took part into the interview, etc) as OSPs, making these estimates directly comparable because subject to the same sources of bias. Nonetheless, our results, as it is the case in most observational studies should be generalised with care.

## 4 Results

Descriptive results show differences in the proportion of smokers by education. Among higher educated people, the percentage of those who have never smoked is considerably higher than among lower educated individuals (50 and 39%, respectively; Table 1). Moreover, SSPs and OSPs vary across education and smoking categories. Table 2a shows that within all three groups defined by smoking behaviours, higher educated people report, on average, about 10 percentage points higher SSPs than the lower educated ($p<0.05$).

In terms of smoking behaviour, never smokers have the highest SSP (60.9), followed by past smokers (59.7) and current smokers (55.3). A similar pattern is observed for both higher and lower educated people. However, among the lower educated the difference between average SSPs of current and of never smokers is not statistically significant ($p>0.05$).

Also the prevalence of deaths observed during the study period (Table 2b) differs by education and smoking behaviour. It is lower for higher educated people than for their lower educated counterpart. Both current and past smokers exhibit higher prevalence of deaths than never smokers ($p<0.05$); while past smokers have a higher prevalence of deaths than current smokers. This unadjusted difference may be due to different age distributions between the two groups (Table 1) and to the fact that past smokers may have quitted smoking because of health reasons.

In the first step of our multivariate analyses we investigate how SSPs vary by sub-groups, i.e., by education and smoking behaviours (never smoked, past smoker, currently smoking), adjusting for the control variables previously described. Figure 1a shows the predicted SSPs for these sub-groups with 95% confidence intervals. Complete estimates of all regression models are available in Tables S.1-S.4 (supplementary materials).

For each smoking status, Fig 1a indicates that the higher educated report significantly higher SSPs than their lower educated counterparts. Within the education sub-groups, respondents who never smoked report a significantly higher SSP than their counterparts who are currently smoking. However, while past smokers report the highest SSP among the lower educated, past smokers and never smokers do not significantly differ in terms of SSP among the higher educated.

**Table 1. Respondents' characteristics in the working sample.**

| | Low Education | | | | | | | | High Education | | | | | | | | Total | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Smoking | | | | | | | | Smoking | | | | | | | | | |
| | Never | | Past | | Current | | Total | | Never | | Past | | Current | | Total | | | |
| **Age** | | | | | | | | | | | | | | | | | | |
| Median (IQR) | 62 | (55-72) | 63 | (56-72) | 57 | (54-63) | 61 | (55-70) | 57 | (53-65) | 61 | (55-69) | 57 | (53-64) | 58 | (54-67) | 61 | (55-69) |
| **Gender** | | | | | | | | | | | | | | | | | | |
| Male | 1954 | (27%) | 3901 | (52%) | 1832 | (46%) | 7687 | (41%) | 1187 | (47%) | 1210 | (58%) | 259 | (52%) | 2656 | (52%) | 10343 | (43%) |
| Female | 5344 | (73%) | 3604 | (48%) | 2135 | (54%) | 11083 | (59%) | 1362 | (53%) | 867 | (42%) | 240 | (48%) | 2469 | (48%) | 13552 | (57%) |
| **Ethnicity** | | | | | | | | | | | | | | | | | | |
| White/Caucasian | 5432 | (74%) | 5869 | (78%) | 2701 | (68%) | 14002 | (75%) | 2087 | (82%) | 1765 | (85%) | 388 | (78%) | 4240 | (83%) | 18242 | (76%) |
| Black/African American | 1347 | (18%) | 1192 | (16%) | 982 | (25%) | 3521 | (19%) | 289 | (11%) | 219 | (11%) | 86 | (17%) | 594 | (12%) | 4115 | (17%) |
| Other | 519 | (7%) | 444 | (6%) | 284 | (7%) | 1247 | (7%) | 173 | (7%) | 93 | (4%) | 25 | (5%) | 291 | (6%) | 1538 | (6%) |
| **Diagnosed illness** | | | | | | | | | | | | | | | | | | |
| No | 5216 | (71%) | 4486 | (60%) | 2588 | (65%) | 12290 | (65%) | 2026 | (79%) | 1454 | (70%) | 355 | (71%) | 3835 | (75%) | 16125 | (67%) |
| Yes | 2082 | (29%) | 3019 | (40%) | 1379 | (35%) | 6480 | (35%) | 523 | (21%) | 623 | (30%) | 144 | (29%) | 1290 | (25%) | 7770 | (33%) |
| **Total** | 7298 | (39%)* | 7505 | (40%)* | 3967 | (21%)* | 18770 | (79%) | 2549 | (50%)* | 2077 | (41%)* | 499 | (10%)* | 5125 | (21%) | 23895 | |

Note: Data are n (%) unless stated otherwise. IQR = Inter-quartile range. * Percentages calculated within educational groups.

Table 2: Average subjective probability to survive to the target age and prevalence of deaths during the observation period by education and smoking.

| Smoking | Education: Low | Education: High | Total |
|---|---|---|---|
| **a)** | **Subjective probability to survive to the target age** | | |
| **Never** | | | |
| Mean | 58.1 | 68.8 | 60.9 |
| 95% CI | (57.3-59.0) | (67.7-69.9)* | (60.2-61.6) |
| n | 7298 | 2549 | 9847 |
| **Past** | | | |
| Mean | 57.8 | 66.4 | 59.7 |
| 95% CI | (56.9-58.6) | (65.1-67.8)*‡ | (58.9-60.4)‡ |
| n | 7505 | 2077 | 9582 |
| **Current** | | | |
| Mean | 54.8 | 59.7 | 55.3 |
| 95% CI | (53.6-56.0)† | (56.7-62.7)*† | (54.2-56.5)† |
| n | 3967 | 499 | 4466 |
| **Total** | | | |
| Mean | 57.3 | 67.0 | 59.4 |
| 95% CI | (56.8-57.8) | (66.1-67.8)* | (58.9-59.8) |
| n | 18770 | 5125 | 23895 |
| **b)** | **Prevalence of deaths during the observation period** | | |
| **Never** | | | |
| % | 19.9 | 11.0 | 17.6 |
| 95% CI | (18.8-20.9) | (9.6-12.4)* | (16.7-18.4) |
| **Past** | | | |
| % | 28.4 | 18.8 | 26.3 |
| 95% CI | (27.2-29.5)‡ | (16.8-20.7)*‡ | (25.3-27.3)‡ |
| **Current** | | | |
| % | 24.1 | 21.3 | 23.8 |
| 95% CI | (22.6-25.7)† | (17.2-25.4)† | (22.4-25.3)† |
| **Total** | | | |
| % | 24.2 | 15.1 | 22.2 |
| 95% CI | (23.5-24.9) | (14.0-16.3)* | (21.6-22.8) |

Note: Data are mean, 95% CI, sample sizes (n) or prevalence (%). *$p<0.05$ for the difference between high and low education. † $p<0.05$ for the difference between current smoker and never smoked. ‡$p<0.05$ for the difference between past smoker and never smoked.

We then calculate, for each respondent, the “gap” between SSP and the estimated OSP from the HRS data. Figure 1b shows the predictions from a regression model where the logarithm of the absolute gap is a function of our explanatory and control variables. The higher the value on the vertical axis, the larger the absolute gap between SSP and OSP, reflecting a higher “mistake” of respondents in guessing their survival probability. For all smoking statuses, higher educated respondents have a smaller gap between SSP and OSP than their lower educated counterparts, meaning that the lower educated are more likely to make mistakes in predicting their chances to survive. The patterns by smoking behaviour within the education sub-groups show that current smokers are the most likely to mistake their expectations of survival. Past smokers are the best at predicting their survival probabilities. However, only among the lower educated their SSPs are significantly closer to their OSPs than it is for their counterparts who never smoked.

These results are only informative of the amount of the (average) gap by smoking and education as they do not distinguish between over- and under-estimates of the survival probabilities. Therefore, in the multinomial analyses shown in Figure 2, the outcome is an unordered categorical variable that distinguishes between overestimation and underestimation of survival probabilities. The reference category is represented by an approximately correct subjective estimation of survival probability defined as a difference between SSP and OSP of maximum ten percentage points in absolute terms. We first look at the predicted probability of estimating own survival chances correctly (Figure 2a). For each smoking status, the higher educated are more likely to correctly predict their survival probabilities than their lower educated counterparts. Within education groups, those who have never smoked are better at predicting their survival probability as compared to current smokers. Yet, among the lower educated, past smokers are significantly better at predicting their survival probability.

Among respondents who “make mistakes” in their SSP, interestingly those who never smoked are the most likely to underestimate their survival probabilities (Figure 2b) while current smokers tend to overestimate them (Figure 2c). Past smokers hold an in-between position in both cases.

## 5 Discussion

This is the first study that jointly analyses the effect of smoking behaviour and education on SSPs and on the ability of survey respondents to predict their real survival.

Using longitudinal data from the Health and Retirement Study, we first compared SSPs of population sub-groups by education and smoking behaviours. People currently smoking reported lower subjective survival probabilities (SSPs) especially if they were lower educated. This is consistent with what is observed in objective survival probabilities (OSPs), which, in a second step of our analysis, were predicted using a Gompertz model on real survival data as a function of educational attainment and smoking behaviour. Third, by comparing the gap between SSPs and OSPs across different population sub-groups, we found that irrespectively of the smoking status, higher educated people were more likely to correctly predict their survival probabilities as compared to their lower educated counterparts. Within education groups, past smokers were the best at predicting their survivorship. Interestingly, current smokers reported the highest probability to overestimate their survival probability, indicating that smokers tend to underestimate the negative effect of smoking on survival.

Previous literature on health-related behaviours has demonstrated that smoking negatively impacts on health and survival (Jayes et al. 2016; Taghizadeh et al. 2016). Other strands of research have also highlighted education effects on health, health behaviours and life expectancy (Cacciani et al. 2015; Bijwaard et al. 2015; Brunello et al. 2016). We found that higher educated people are more aware of the risks of smoking: high education is associated with better prediction of own survivorship. We also showed that smoking and education play together in determining how well people can assess the own survival “potential”.

One limitation of our study is that we could not differentiate smokers according to the number of cigarettes smoked and past smokers according to when they stopped smoking, which might moderate the health consequences (Hoogenveen et al. 2008). Future studies on SSPs may distinguish these sub-groups. Moreover, an interesting avenue for extending this study could be to investigate to what extent high educated smokers, being aware of the risks of smoking, compensate them by reducing other risky behaviours, such as drinking and physical inactivity. Further research is also needed on how the intersection of various risky behaviours and socio-economic characteristics modify individuals’ perceptions of their survival chances.

Policy makers can draw some relevant conclusions from our study to design policies concerned with health and survivorship in later life. Our results help shedding light on whether and to what extent individuals understand the mortality consequences of smoking and on the role of education in increasing people awareness. Despite the various anti-smoking campaigns and smoking restrictions, smokers may not be fully aware of the risks of smoking. In particular, educational groups seem to be differently exposed to the information disseminated. Our findings suggest the need to target the dissemination of such information to lower educated people that are the most likely to underestimate the risks of smoking. Providing information on survival probabilities by smoking behaviour may not only reduce smoking but it may also increase individuals’ ability to assess their own survival.

The fact that sub-groups within the population differently incorporate the effects of smoking into their assessment of survival probabilities may have important consequences for decisions in different spheres of (later) life: retirement, investments, and healthy behaviours(Hamermesh 1985; Salm 2010; Carbone et al. 2005; Scott-Sheldon et al. 2010; Adams et al. 2015). Our study, in line with previous literature suggesting that SSPs are a good (subjective) measure of health, provides wide scope for future research based on SSPs.

Figure 1. a) Predicted subjective survival probability by education and smoking behaviour; b) Predicted logarithm of the gap between subjective and objective survival probabilities.

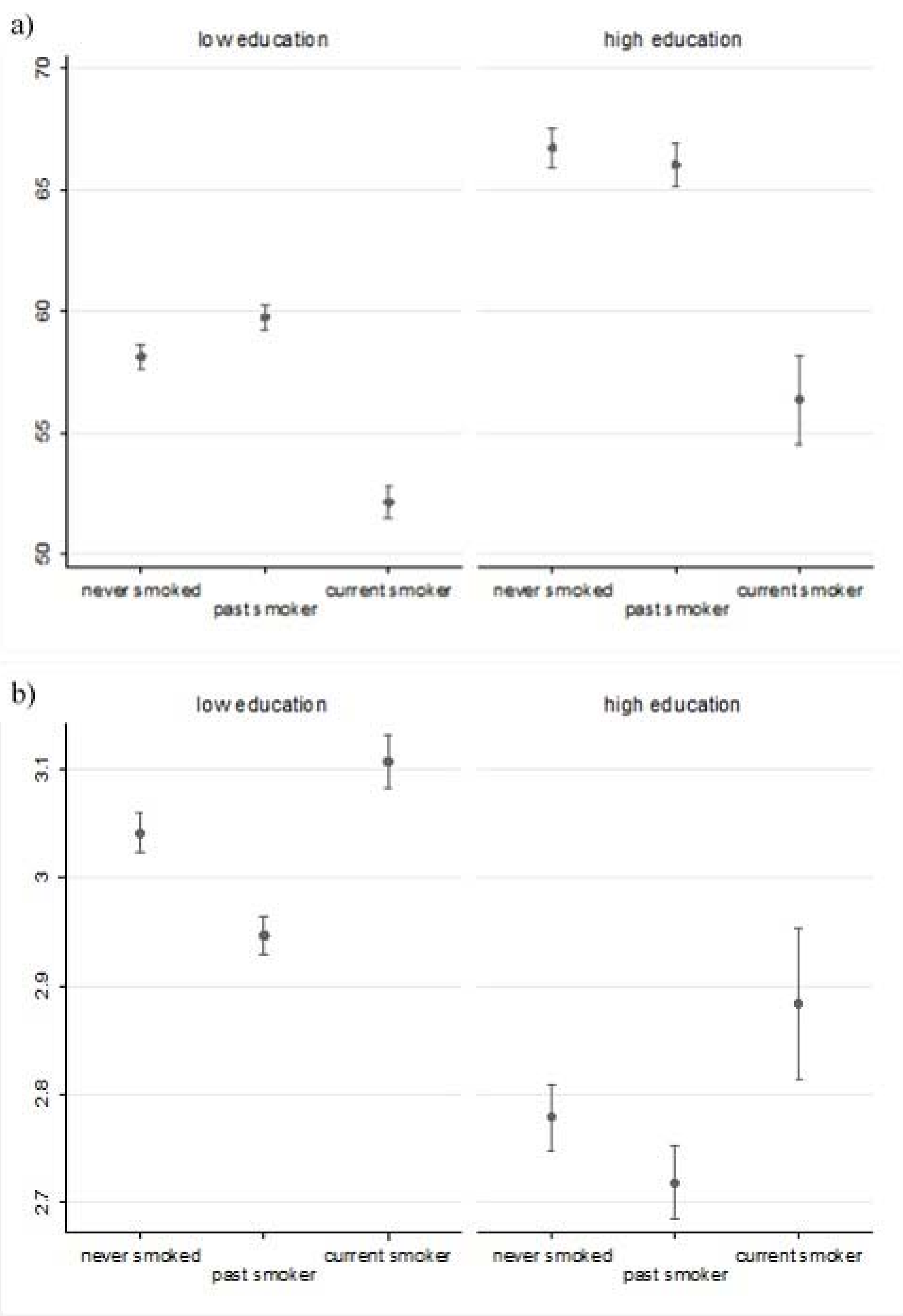
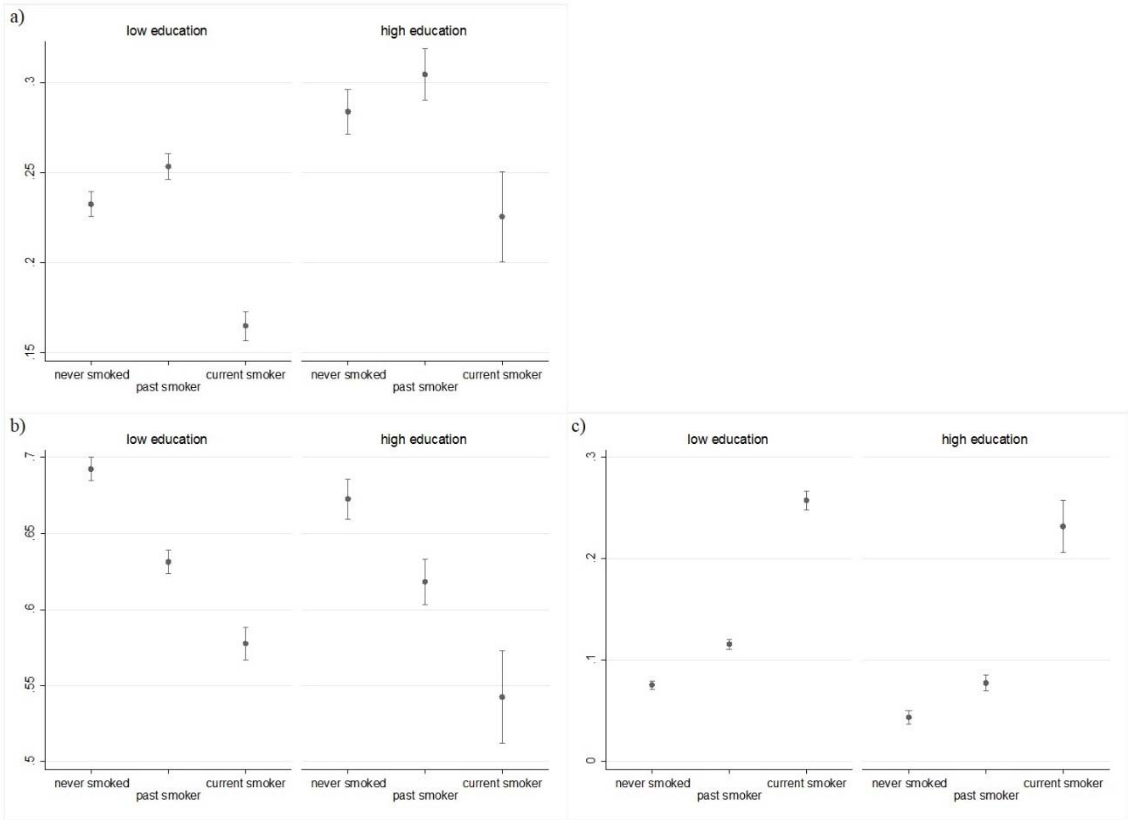


Figure 2. a) Probabilities of correctly estimate the own survival probabilities; b) Probabilities of underestimating the own survival probabilities; c) Probabilities of overestimating the own survival probabilities.




## 6 References

Adams, J. et al. 2015. Anticipated survival and health behaviours in older English adults: Cross sectional and longitudinal analysis of the English Longitudinal Study of Ageing. *PLOS ONE* 10(3): e0118782.

Balia, S. 2014. Survival expectations, subjective health and smoking: evidence from SHARE. *Empirical Economics* 47(2): 753–780.

Bijwaard, G.E. et al. 2015. Gains in life expectancy associated with higher education in men. *PLOS ONE* 10(10): e0141200.

Bissonnette, L., M.D. Hurd, and P.-C. Michaud. 2014. *Individual Survival Curves Combining Subjective and Actual Mortality Risks*. Bonn, Germany: Institute for the Study of Labor (IZA).

Brunello, G. et al. 2016. The causal effect of education on health: What is the role of health behaviors? *Health Economics* 25(3): 314–336.

Cacciani, L. et al. 2015. Education and mortality in the Rome Longitudinal Study. *PLOS ONE* 10(9): e0137576.

Carbone, J.C., S. Kverndokk, and O.J. Røgeberg. 2005. Smoking, health, risk, and perception. *Journal of Health Economics* 24(4): 631–653.

Dominitz, J. and C.F. Manski. 1997. Using expectations data to study subjective income expectations. *Journal of the American Statistical Association* 92(439): 855–867.

van Doorn, C. and S.V. Kasl. 1998. Can parental longevity and self-rated life expectancy predict mortality among older persons? Results from an Australian cohort. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences* 53(1): S28-34.

Elder, T. 2013. The predictive validity of subjective mortality expectations: Evidence from the Health and Retirement Study. *Demography* 50(2): 569–589.

Falba, T.A. and S.H. Busch. 2005. Survival expectations of the obese: Is excess mortality reflected in perceptions? *Obesity Research* 13(4): 754–761.

Gilman, S.E. et al. 2008. Educational attainment and cigarette smoking: a causal association? *International Journal of Epidemiology* 37(3): 615–624.

Hamermesh, D.S. 1985. Expectations, life expectancy, and economic behavior. *The Quarterly Journal of Economics* 100(2): 389–408.

Hoogenveen, R.T. et al. 2008. Dynamic effects of smoking cessation on disease incidence, mortality and quality of life: The role of time since cessation. *Cost Effectiveness and Resource Allocation: C/E* 6: 1.

Hurd, M.D. 2009. Subjective probabilities in household surveys. *Annual Review of Economics* 1(1): 543–562.

Hurd, M.D. and K. McGarry. 1995. Evaluation of the subjective probabilities of survival in the health and retirement study. *The Journal of Human Resources* 30: S268–S292.

Hurd, M.D. and K. McGarry. 2002. The predictive validity of subjective probabilities of survival. *The Economic Journal* 112(482): 966–985.

Jayes, L. et al. 2016. SmokeHaz: Systematic reviews and meta-analyses of the effects of smoking on respiratory health. *Chest* 150(1): 164–179.

Khwaja, A., F. Sloan, and S. Chung. 2007. The relationship between individual expectations and behaviors: Mortality expectations and smoking decisions. *Journal of Risk and Uncertainty* 35(2): 179–201.

Kutlu-Koc, V. and A. Kalwij. 2017. Individual survival expectations and actual mortality: Evidence from Dutch survey and administrative data. *European Journal of Population*: 1–24.

Ludwig, A. and A. Zimper. 2013. A parsimonious model of subjective life expectancy. *Theory and Decision* 75(4): 519–541.

Manski, C.F. 2004. Measuring expectations. *Econometrica* 72(5): 1329–1376.

Novak, B. and A. Palloni. 2013. *Subjective Survival Expectations and Observed Survival: How Consistent Are They?*. University of Wisconsin-Madison: Center for Demography and Ecology.

Perozek, M. 2008. Using subjective expectations to forecast longevity: do survey respondents know something we don't know? *Demography* 45(1): 95–113.

Salm, M. 2010. Subjective mortality expectations and consumption and saving behaviours among the elderly. *Canadian Journal of Economics/Revue canadienne d'économique* 43(3): 1040–1057.

Scott-Sheldon, L.A.J. et al. 2010. Subjective life expectancy and health behaviors among STD clinic patients. *American Journal of Health Behavior* 34(3): 349–361.

Siegel, M., E.H. Bradley, and S.V. Kasl. 2003. Self-rated life expectancy as a predictor of mortality: evidence from the HRS and AHEAD surveys. *Gerontology* 49(4): 265–271.

Smith, V.K., D.H. Taylor, and F.A. Sloan. 2001. Longevity expectations and death: Can people predict their own demise? *American Economic Review* 91(4): 1126–1134.

Taghizadeh, N., J.M. Vonk, and H.M. Boezen. 2016. Lifetime smoking history and cause-specific mortality in a cohort study with 43 years of follow-up. *PLOS ONE* 11(4): e0153310.

Wang, H. and S.H. Preston. 2009. Forecasting United States mortality using cohort smoking histories. *Proceedings of the National Academy of Sciences of the United States of America* 106(2): 393–398.

Wang, Y. 2014. Dynamic implications of subjective expectations: Evidence from adult smokers. *American Economic Journal: Applied Economics* 6(1): 1–37.

Wilson, D.L. 1994. The analysis of survival (mortality) data: fitting Gompertz, Weibull, and logistic functions. *Mechanisms of Ageing and Development* 74(1–2): 15–33.

## Appendix

Table A1. Adjusted association between education, smoking and subjective survival probabilities estimated using a linear regression model.

| Independent variables | b | 95 CI % |
|---|---|---|
| Education: (reference: low education) | | |
| High education | 8.60*** | (7.26, 9.94) |
| Smoking: (reference: never smoked) | | |
| Past smoker | 1.61*** | (0.64, 2.58) |
| Current smoker | -5.99*** | (-7.15, -4.83 |
| Interactions education x smoking: | | |
| High education x Past smoker | -2.31** | (-4.25, -0.36) |
| High education x Current smoker | -4.39*** | (-7.43, -1.36) |
| Age | 0.02*** | (0.01, 0.03) |
| Gender (reference: male) | | |
| Female | 2.12*** | (1.34, 2.89) |
| Ethnicity (reference: White/Caucasian) | | |
| Black/African American | 3.98*** | (2.94, 5.01) |
| Other | -6.08*** | (-7.65, -4.52) |
| Diagnosed illness (reference: no) | | |
| Yes | -10.47*** | (-11.30, -9.65) |
| Target (reference: target = 75) | | |
| 80 | -15.71*** | (-20.57, -10.86) |
| 85 | -6.96* | (-14.55, 0.63) |
| 90 | 5.44 | (-7.54, 18.41) |
| 95 | -18.61* | (-40.59, 3.36) |
| 100 | -15.41 | (-54.67, 23.85) |
| Interactions age x target: | | |
| age x 80 | 0.12*** | (0.04, 0.21) |
| age x 85 | -0.05 | (-0.12, 0.02) |
| age x 90 | -0.18*** | (-0.25, -0.10) |
| age x 95 | -0.06 | (-0.16, 0.04) |
| age x 100 | -0.08 | (-0.22, 0.05) |
| Constant | 67.44*** | (66.25, 68.64) |
| N | 23,895 | |

Note: Estimates from a linear regression model where the outcome is the subjective survival probability. b = unstandardized coefficients. *** p<0.01; ** p<0.05; * p<0.1.

Table A2. Adjusted association between education, smoking and (objective) survival estimated using a Gompertz survival model.

| Independent variables | b | 95 CI % |
|---|---|---|
| Education: (reference: low education) | | |
| High education | -0.27*** | (-0.40, -0.13) |
| Smoking: (reference: never smoked) | | |
| Past smoker | 0.29*** | (0.22, 0.36) |
| Current smoker | 0.92*** | (0.84, 1.01) |
| Interactions education x smoking: | | |
| High education x Past smoker | -0.00 | (-0.17, 0.17) |
| High education x Current smoker | 0.18 | (-0.06, 0.42) |
| Age | 0.01*** | (0.01, 0.01) |
| Gender (reference: male) | | |
| Female | -0.24*** | (-0.30, -0.19) |
| Ethnicity (reference: White/Caucasian) | | |
| Black/African American | 0.24*** | (0.16, 0.32) |
| Other | -0.04 | (-0.21, 0.14) |
| Diagnosed illness (reference: no) | | |
| Yes | 0.59*** | (0.54, 0.65) |
| Constant | -13.66*** | (-13.92, -13.39) |
| Gamma | 0.01*** | (0.01, 0.01) |
| N | 23,895 | |

Note: Estimates from a Gompertz survival model where the outcome is objective survival. b = unstandardized coefficients. *** p<0.01; ** p<0.05; * p<0.1. Gamma is the estimate of the ancillary parameter of the Gompterz distribution.

Table A3. Adjusted association between education, smoking and the gap between subjective and objective survival probabilities estimated using a linear regression model.

| Independent variables | b | 95 CI % |
|---|---|---|
| Education: (reference: low education) | | |
| High education | -0.26*** | (-0.31, -0.21) |
| Smoking: (reference: never smoked) | | |
| Past smoker | -0.09*** | (-0.13, -0.06) |
| Current smoker | 0.07*** | (0.02, 0.11) |
| Interactions education x smoking: | | |
| High education x Past smoker | 0.03 | (-0.04, 0.11) |
| High education x Current smoker | 0.04 | (-0.08, 0.16) |
| Age | 0.05*** | (0.04, 0.07) |
| Gender (reference: male) | | |
| Female | -0.06*** | (-0.09, -0.03) |
| Ethnicity (reference: White/Caucasian) | | |
| Black/African American | -0.05*** | (-0.09, -0.01) |
| Other | 0.23*** | (0.17, 0.29) |
| Diagnosed illness (reference: no) | | |
| Yes | 0.14*** | (0.11, 0.17) |
| Constant | 2.60*** | (2.47, 2.73) |
| N | 23,895 | |

Note: Estimates from a linear regression model where the outcome is the logarithm of the absolute difference between subjective and objective survival probabilities. b = unstandardized coefficients. *** p<0.01; ** p<0.05; * p<0.1.

Table A4. Adjusted association between education, smoking and the probability to underestimate or overestimate subjective survival probabilities estimated using a multinomial logistic model.

| Independent variables | Underestimated versus correct | | Overestimated versus correct | |
|---|---|---|---|---|
| | b | 95 CI % | b | 95 CI % |
| Education: (reference: low education) | | | | |
| High education | -0.23*** | (-0.34, -0.13) | -0.80*** | (-1.07, -0.54) |
| Smoking: (reference: never smoked) | | | | |
| Past smoker | -0.18*** | (-0.26, -0.10) | 0.40*** | (0.27, 0.53) |
| Current smoker | 0.18*** | (0.07, 0.28) | 1.81*** | (1.66, 1.96) |
| Interactions education x smoking: | | | | |
| High education x Past smoker | 0.02 | (-0.13, 0.17) | 0.16 | (-0.17, 0.49) |
| High education x Current smoker | -0.15 | (-0.40, 0.11) | 0.35* | (-0.05, 0.76) |
| Age | 0.00*** | (0.00, 0.00) | 0.01*** | (0.01, 0.01) |
| Gender (reference: male) | | | | |
| Female | -0.11*** | (-0.17, -0.05) | -0.34*** | (-0.44, -0.24) |
| Ethnicity (reference: White/Caucasian) | | | | |
| Black/African American | -0.27*** | (-0.35, -0.19) | 0.83*** | (0.71, 0.95) |
| Other | 0.24*** | (0.11, 0.37) | 0.36*** | (0.13, 0.60) |
| Diagnosed illness (reference: no) | | | | |
| Yes | 0.26*** | (0.19, 0.33) | 0.85*** | (0.75, 0.96) |
| Constant | 0.66*** | (0.42, 0.89) | -7.62*** | (-8.02, -7.22) |
| N | 23,895 | | | |

Note: Estimates from a multinomial logistic regression model where the outcome is a categorical variable indicating that the subjective survival probability indicated by the respondent is: correct (reference), underestimated or overestimated with respect to his/her objective survival probability. Correct is defined as an absolute difference between subjective and objective survival probabilities not bigger than 10 percentage points. b = unstandardized coefficients. *** p<0.01; ** p<0.05; * p<0.1.