Network for Studies on Pensions, Aging and Retirement

Netspar DISCUSSION PAPERS

Vesile Kutlu-Koc and Adriaan Kalwij

Individuals' Survival Expectations and Actual Mortality

DP 05/2013-013

Individuals' Survival Expectations and Actual Mortality

Vesile Kutlu-Koc,[a] Adriaan Kalwij[b]

May 6, 2013

Abstract

Using a combination of Dutch survey and administrative data, we show that survival expectations do in fact predict actual mortality in models that control for income and education level. This predictive power disappears, however, when controls are introduced for self-rated health status and smoking behavior. Concerning the differences between survival expectations and actual mortality, our results show that, on average, women underestimate their remaining life duration more than men and that the age gradient is steeper in subjective than in actual mortality. The association of current health status with subjective survival is less strong than with objective survival and, moreover, individuals underestimate the risks from smoking, alcohol consumption, and obesity.

[a] Corresponding author. V.Kutlu@uu.nl, Utrecht University School of Economics, Network for Studies on Pensions, Aging and Retirement (Netspar).

[b] A.S.Kalwij@uu.nl, Utrecht University School of Economics, Tilburg University, Network for Studies on Pensions, Aging and Retirement (Netspar).

1. Introduction

Individual life expectancy plays an important role in life cycle models of economic behavior. Several household surveys, therefore, include probabilistic questions (Manski 2004) to collect information on individuals' beliefs about their survival chances to one or two target ages (Hurd 2009). Most particularly, the existing literature indicates that these subjective survival probabilities (SSPs) elicited from survey respondents convey useful information about their actual life duration. This observation has an important practical implication for economic research. If SSPs are accurate predictors of individuals' actual survival, they can be used as measures of individual mortality risk in economic models, including life-cycle models of saving, consumption, and retirement behavior.[1,2] Yet, earlier research indicates that life-cycle models, which replace individual mortality risk with that from (actuarial) life tables, cannot explain certain well-known anomalies in the data (Hurd and McGarry 2002). SSPs may shed light on this issue as they convey information on individuals' beliefs about survival, which can differ from actuarial survival probabilities. For instance, the anomaly of insufficient preretirement savings to finance retirement consumption can in part be explained by models that use SSPs. If prior to retirement, individuals systematically underestimate their remaining life duration, hence their beliefs about survival are lower than their actual survival to a certain age, their savings level will be lower than that implied by a life-cycle model of consumption based on actuarial survival probabilities. In fact, Gan et al. (2004) and Salm (2010) show that individuals have private

[1] Individual mortality risk is the risk associated with the uncertainty of an individual's remaining lifetime conditional on given survival probabilities (De Waegenaere et al. 2010).

[2] Also, in this case, studies investigating the socio-economic gradient in mortality may use SSPs as substitutes for actual mortality data (Delavande and Rohwedder 2011).

information about their own mortality risk on which they base their economic decisions. Our paper focuses on the step that needs to be taken before developing and estimating economic models that incorporate SSPs, and explores the extent to which individuals' beliefs about survival (SSPs) relate to their actual mortality and how this differs with socioeconomic status and health characteristics.

The contribution to the literature of our paper is threefold. First, we investigate whether (remaining) life expectancy of the Dutch measured by their SSPs correlate with actual mortality risk and assess its predictive power. Second, by estimating subjective and objective mortality risk models, we analyze the extent to which individual SSPs contain well-known differences in mortality risk with respect to socioeconomic and health status, and behavioral risk indicators such as smoking and drinking. Finally, we use the estimation results to quantify the gap between individuals' subjective life expectancy and the life duration implied by observed mortality for different types of individuals. For this purpose, we use data from the DNB Household Survey (DHS) supplemented with administrative data on actual mortality. We compute subjective remaining lifetime using two SSPs for each individual under minimal assumptions. Methodologically, therefore, we are able to estimate the same parameters of a mortality risk model using both subjective and objective survival information.

The empirical literature on this topic can be divided into two categories: The first investigates the link between SSPs and actuarial survival probabilities (Hamermesh 1985; Hurd and McGarry 1995; O'Donnell et al. 2008; Perozek 2008; Peracchi and Perotti 2011; Teppa 2012). The second assesses the relation between SSPs and actual (objective) mortality and whether SSPs can predict actual mortality within the sample (Van Doorn and Kasl 1998; Smith et al. 2001; Hurd and McGarry 2002; Siegel et al. 2003; Peracchi and Perotti 2011). The research findings in the first stream suggest that males overestimate while females underestimate their survival probabilities compared to life table survival probabilities, although on average, males tend to assess their life expectancy better than females. Nonetheless, Teppa (2012), using Dutch data, finds that both Dutch males and females have lower SSPs relative to actuarial survival probabilities, on average. The main finding of the second stream is that, on average, individuals who expect to live longer are less likely to die. Hurd and McGarry (2002), for instance, report that an increase in the subjective survival probability from 0 to 1 reduces the mortality rate by 53 percent.[3]

Most of the above studies, however, tend to estimate models in which actual mortality is explained by the subjective probability of survival up to age 75, together with such other mortality determinants as income, wealth, education, smoking, and health indicators. One notable exception is Perozek (2008), a study closely related to ours, which fits subjective survival functions using the two SSPs available for each respondent in the HRS survey to generate subjective cohort tables for males and females.[4] In contrast with Perozek (2008), we use subjective survival information to estimate the parameters of a mortality risk model that we also estimate using actual mortality information. We are thus able to assess whether the socioeconomic status and health-related parameters estimated using the subjective data are close to those obtained using actual mortality data.

[3] See also Van Doorn and Kasl (1998) and Siegel et al. (2003).

[4] This U.S. study investigates whether the Social Security Actuary's (SSA) revision to the gender gap in longevity can be predicted by the gender gap implied by the subjective cohort life tables. Its main finding is that subjective cohort tables predict a smaller difference in life expectancy between men and women than do the SSA predictions. This relatively small gender gap in longevity results from the systematic overestimation of the mortality risk by women relative to men.

A second study that is directly linked to ours is Delavande and Rohwedder (2011), which compares the coefficient estimates of an actual survival model with those implied within the same sample by a subjective survival model. The findings suggest that Americans' actual mortality and subjective survival expectations are similarly associated with wealth, income, and education, which implies that in the case of the United States, SSPs are informative proxies for actual survival. This study, however, examines survival only up to age 75, whereas our sample includes respondents aged 75 or older. More importantly, when comparing actual survival with subjective survival, we control for socioeconomic variables as well as health indicators. Furthermore, unlike Delavande and Rohwedder (2011) which adds age dummies to control for respondents' baseline age, we use two SSPs for each respondent which, together with the assumption of a Gompertz hazard function, allows us to estimate the age gradient in subjective mortality risk.

One major finding of our paper is that when we control for income and education level, SSPs predict observed mortality within the sample, but when we control for self-rated health status and smoking behavior, the correlation disappears. This outcome may suggest that SSPs' predictive power—being largely determined by the (observed) current health situation with little additional information on expected future health events—is rather limited. We also show that, men and women underestimate, on average, their remaining life duration by about 1 and 8 years, respectively, meaning that their long-term decisions on such factors as retirement and savings may be suboptimal. The association of current health status with subjective survival is less strong than with objective survival and individuals underestimate the risks from smoking, obesity, and alcohol consumption.

The paper is structured as follows: Section 2 describes the data. Section 3 outlines the mortality risk model and the methods for its estimation using objective (actual) and then subjective survival data. Section 4 presents the estimation results, and Section 5 offers our concluding remarks.

2. Data

To compare Dutch individuals' survival expectations with their actual mortality (i.e., date of death), this paper combines survey data with individual-level administrative data. The SSPs (subjective survival probabilities) are taken from the 1995 and 1996 waves of the DNB Household Survey (DHS), begun in 1993 and originally known as the CentER Savings Survey. The DHS database, compiled using an Internet survey of around 2,550 Dutch households, includes detailed information on respondents' age, income, health, education, labor market status, assets and liabilities, and psychological state (see Alessie et al. 2002 for a detailed description). All of the DHS questions are asked of two different panels, one a nationwide representative panel of around 1,900 households and the other, a high-income panel of around 650 households that represent the top 10 percent of the income distribution. Every year, all household members aged 16 or over are interviewed online. Those who do not have a computer and/or Internet access are provided with these tools by the survey agency.

The data on the actual mortality of survey respondents are obtained from the Dutch causes of death registry (DO, DoodsOorzaken) survey, which records the date of death of all residents deceased during the 1995–2010 period. These data are provided by medical examiners, who are legally obliged to submit them to Statistics Netherlands. The DO dataset also assigns a personal identifier that matches the personal identifier in the DHS, thereby allowing determination of whether individuals in the 1995 or 1996 wave of the DHS were still alive at the end of the observation period (December 31, 2010) or whether they had died, and if so, on which date.

The DHS measures subjective survival probabilities using the following survey question:

- *How big do you think is the chance that you will attain (at least) the age of T?*

where $T \in \{75, 80, 85, 90, 95, 100\}$ is a target age that depends on the respondent's current age.[5] Respondents aged 16 through 65 report their probability of survival to age 75; those aged 16 through 70, their survival expectations to age 80; and respondents aged 65–75, 70–80, 75–85, and 80–90, their expected survival probabilities to 85, 90, 95, and 100, respectively (see Table 1). The responses are measured on a 10-point scale, from 0, "no chance at all," to 10, "absolutely certain." Following Hurd and McGarry (1995), we assume that after being divided by 10, the responses can be interpreted as probabilities conditional on being alive at a certain age.[6] To construct our main variable of interest, median remaining life duration, we use two survival probabilities for each individual (see Section 2.2.1).

[5] The framing of the question may affect respondents' answers to SSP. For instance, respondents might have provided different answers if they had been asked "*How big do you think is the chance that you will attain (at most) the age of T*?" Whereas the actual survey question asks individuals' probability of living to a certain age or older, this latter asks their chance of dying by a certain age and younger. Payne et al. (2013) find that individuals report higher SSPs in the first than in the second.

[6] For computational reasons (see Appendix A), we replace probabilities 0 and 1 with 0.01 and 0.99, respectively.

2.1. Sample Selection

We include in our sample individuals aged 25 and over.[7] The DHS dataset, which is cross-sectional, consists of one observation per individual in either 1995 or 1996. If respondents were observed in both the 1995 and 1996 waves, we use only the earlier response to avoid the potential influence of repeated interviewing on respondent behavior (Lazarsfeld 1940; Sturgis et al. 2009); for example, respondents asked about survival probability in 1995 may seek more information about their survival chances before responding in 1996. Our method eliminates the risk of possible learning effects.

Manski (2004) suggests that "showing that respondents are willing and able to respond to probabilistic questions is an obvious prerequisite for substantive interpretation of the data" (p. 1342). Table 1 lists the SSP response rates, which in our sample, total an average of 86.35 percent, a considerably lower response rate than in the HRS (about 98 percent) or SHARE (about 90 percent) surveys (Hurd and McGarry 1995; Peracchi and Perotti 2011).

A relatively easy way to assess respondents' understanding of the SSP questions is to check whether they can provide internally consistent answers. For example, if a respondent says that his survival probability to age 75 is less than or equal to his survival probability to age 80, his answer violates the strict monotonicity assumption. In fact, his survival probability to age 75 should be greater than his survival probability to age 80 because to reach age 80, he must first survive until age 75. There is also a risk of mortality between these two ages.

[7] Because many individuals under 25 are still enrolled in education, their individual income and (final) educational level are unavailable. We run a robustness test by using a different age group (a sample of individuals older than 50) and our main results are to a large extent unchanged (see Appendix D).

As Table 2 shows, about 67 percent of respondents gave answers that satisfy the strict monotonicity assumption. Around 32 percent reported equal survival probabilities for two target ages, while about 0.63 percent indicated a survival chance to the earlier target age that is less than their survival chance to the later target age. Perozek (2008) suggests that respondents with equal survival probabilities can still give valuable information about the shape of individual subjective survivor functions. Respondents who provide equal survival probabilities for two target ages might, for example, be rounding out their true survival probabilities to the nearest tenth because the answer to survival probability questions ranges from 0 to 10. However, the inclusion of equal survival probabilities in the estimation necessitates arbitrary assumptions and we therefore exclude these 32.73 percent with inconsistent answers from our analysis. Only as a sensitivity check, we have included respondents with equal probabilities in our sample and assume a 5 percent shift between equal probabilities (see Tables C1–C2 in Appendix C). For brevity, we do not detail these additional results, but in general it can be concluded that they are sensitive to the inclusion of equal survival probabilities in the estimation.[8] Finally, there is also a tendency for respondents to provide focal point answers on SSPs (i.e., clustered around 0, 0.5, and 1; see Table B1, Appendix B). In the face of Kleinjans and van Soest's (2013) evidence that taking into account rounding or focal point answers does not substantially change the coefficient estimates on the determinants of subjective probabilities of survival, we make no corresponding corrections to the SSP responses.

[8] We also assign a 10 percent shift between equal probabilities as suggested by Perozek (2008). The estimation results again change slightly compared to those from the 5 percent shift assumption but they are still significantly different than the results without equal probabilities.

In sum, at baseline, the sample includes 5,747 observations and we are forced to drop 2,688 observations (46.8 percent of the initial sample) either because the strict monotonicity assumption for SSP responses is violated or because information is missing for one of the covariates under study. Although we are solely concerned with within sample predictions, we test if individuals in our final sample of 3,059 individuals (10.8 percent of whom were deceased by December 2010) have the same mortality risk as those dropped from the baseline sample. We find no significantly different mortality risk (at $p = 0.102$), indicating that the sample selection is not endogenous with respect to mortality. The construction of our median remaining life duration variable is detailed in the next section; the definitions of our other variables are given in Table B2, Appendix B.

2.2. Descriptive Statistics

2.2.1. Median remaining life duration

Using the respondents' subjective survival probabilities, we compute the subjective median remaining life duration conditional on baseline age for each individual in the sample (equation (A7), Appendix A). As Figure 1 shows, the mean of the median remaining life duration is 53 years for the group aged 25–29 but decreases to about 12 years for the individuals aged 80–84.

We observe the date of death of 330 individuals in our sample that die before December 2010. A comparison of the subjective and objective (within-sample) remaining life durations across age categories and gender (see Table 3) reveals that for females, the subjective values are less than the objective values at all ages, whereas for males, the subjective values are less than the objective values up to age 65 but then exceed their objective counterparts. On average, the difference between subjective and objective remaining life duration is smaller for males than for females. We do note, however, that the remaining life durations based on the HMD life table are always less than their objective counterparts at all ages for both males and females, which may suggest that relatively healthier individuals may be overrepresented in our sample.[9]

As shown in Table 4, the difference between subjective and objective remaining life duration is minimum for males with low education and for females whose self-reported health status is poor. Overall, the statistics in this section indicate that our "subjective median remaining life duration" measure is informative for within-sample mortality.

2.2.2. Control variables

Table 5 presents the mean, median, and standard deviation of the right-hand side variables and baseline age. Women, with a 45 percent share, are slightly underrepresented in our sample compared to men. The mean age at time of interview is about 47 years, 87.6 percent of the respondents are married, 31.2 percent are smokers, 8 percent drink alcohol,[10] 81.6 percent self-rate their health as good or excellent, and 24.5 percent report a chronic condition such as a long term illness, disorder, or disability. Based on the body weight index (BMI), 33.6 percent of the respondents are overweight, and 6 percent are obese. The sample consists mostly of medium educated individuals with pre-university education or junior/senior vocational training, and the average annual standardized household income is ƒ44,287 (€20,096). Standardized household income is defined as the sum of the net annual incomes of all household members divided by the

[9] For this calculation, we use the 1995 life table available in the Human Mortality Database (HMD). Because cohort life tables are not available for each individual in our sample, we are forced to use period life tables, which, when mortality rates decline over time, can tend to underestimate life expectancy compared to cohort life tables.

[10] The alcohol variable is a dummy variable that takes the value of one if the respondent consumes more than four alcoholic drinks a day and zero otherwise.

equivalence scale provided by Statistics Netherlands (Siermann *et al*., 2004). Following Delavande and Rohwedder (2011), we create income terciles, which are insensitive to outliers, by dividing the income distribution into three parts. Although not shown in Table 5, 28 percent of the respondents are members of the high-income panel, while 72 percent fall into the nationwide representative panel.

3. Estimation methodology

Following previous empirical studies on individual mortality, we assume that life duration can be modeled with a (truncated) Gompertz distribution (see, e.g., Gompertz 1825; Olshansky and Carnes 1997; Perozek 2008). One important advantage of this assumption (as demonstrated below) is that it facilitates comparison between the estimated parameters of the subjective and objective mortality models.

3.1. Objective Mortality Model

Assuming that respondent i is aged t_0 when he reports his probability of survival to age t, and T is a random variable representing the respondent's age at death, then the survival function, which gives the probability of the respondents' age at death being greater than t, can be written as follows:

$$S(t|x_i) = \Pr\left(T > t|t_0, \mathbf{x_i}\right) = \exp\left\{-\int_{t_0}^{t} \theta(s|\mathbf{x_i})ds\right\} = \exp\left(-\Lambda_{t_0}(t|\mathbf{x_i})\right), \tag{1}$$

where $\theta(t|\mathbf{x_i})$ is the hazard function of the respondent with characteristics $\mathbf{x_i}$ and $\Lambda_{t_0}(t|\mathbf{x_i})$ is the integrated hazard from age t_0 to age t. Assuming that the random variable T follows a Gompertz distribution, the hazard function can be given by

$$\theta(t|\mathbf{x_i}) = \lambda_i \exp\{\gamma_i t\} \text{ , where } \lambda_i = \lambda_O = \exp\{\mathbf{x_i}\boldsymbol{\beta}_\mathbf{O}\} \text{ and } \gamma_i = \gamma_O > 0 \tag{2}$$

Each respondent is observed first at age $t_{0,i}$; if the respondent dies at t_i, the contribution of this observation to the likelihood function is the density at that duration:

$$L_i(\boldsymbol{\beta}_\mathbf{O}, \gamma_O | \mathbf{x_i}, t_{0,i}, t_i) = f(t_i|\mathbf{x_i}) = S(t_i|\mathbf{x_i})\theta(t_i|\mathbf{x_i}) . \tag{3}$$

If the respondent is still alive at the end of observation period (where t_i is December 2010), the observation is right-censored, and its contribution to the likelihood is

$$L_i(\boldsymbol{\beta}_\mathbf{O}, \gamma_O | \mathbf{x_i}, t_{0,i}, t_i) = S(t_i|\mathbf{x_i}) . \tag{4}$$

By combining equations (3) and (4), we can write the log-likelihood function for the whole sample as

$$\log \prod_{i=1}^{N} L_i(\boldsymbol{\beta}_\mathbf{O}, \gamma_O | \mathbf{x_i}, t_{0,i}, t_i) = \sum_{i=1}^{N} \left(d_i \left(\mathbf{x_i}\boldsymbol{\beta}_\mathbf{O} + \lambda_O t \right) - \Lambda_{t_0}(t|\mathbf{x_i}) \right) \tag{5}$$

where *N* is the number of individuals in our sample and d_i is a dummy variable that takes 1 if the respondent has died at time t_i and 0 otherwise. Based on equation (5), we obtain the maximum likelihood estimates $\hat{\boldsymbol{\beta}}_\mathbf{O}$ and $\hat{\gamma}_O$.

3.2. Subjective Mortality Model

For this model, we use the subjective information on mortality to estimate a set of parameters analogous to the parameters of the objective mortality model. As in the objective mortality model, we assume a Gompertz hazard function given by

$$\theta(t|\mathbf{x_i}) = \lambda_i \exp\{\gamma_i t\} \text{ , where } \lambda_i = \lambda_S = \exp\{\mathbf{x_i}\boldsymbol{\beta}_S\} \text{ and } \gamma_i = \gamma_S > 0$$

We can then estimate the following system of linear equations:

$$\gamma_i^* = \gamma_S + \varepsilon_{1i} \tag{6}$$

$$\ln\left(\lambda_i^*\right) = \mathbf{x}_i^{'}\boldsymbol{\beta}_S + \varepsilon_{2i} \tag{7}$$

where γ_i^* and λ_i^* are the estimated parameters of the individual survival functions (as shown in Appendix A), and ε_{1i} and ε_{2i} stand for the error terms, which are allowed to be correlated with each other. We obtain the $\hat{\boldsymbol{\beta}}_S$ and $\hat{\gamma}_S$ estimates using Seemingly Unrelated Regressions estimation (Zellner, 1962).

4. Empirical Results

4.1. The predictive power of subjective survival for actual mortality

We first estimate a mortality risk model as outlined in Section 3.1 and then stepwise, include a set of covariates (socioeconomic variables and health indicators) in addition to the subjective (median) remaining life duration.[11] Based on a likelihood ratio test, we do not reject pooling the male and female samples at a 5 percent level of significance (for all models) and so include a control variable for gender instead of reporting separate results for men and women.

The estimation results are given in Table 6. The first model estimated explains mortality risk only as a function of year dummy, age, and subjective remaining life duration. The coefficient of subjective remaining life duration is negative and statistically significant, suggesting, in line with the previous studies, that those who expected to live longer at the baseline year experienced a

[11] When we replace the median remaining life duration with its natural logarithm, the results are rather similar; its coefficient is significant at the 5 percent significance level in all models shown in Table 6 except model (4).

lower mortality risk than those whose life expectation was shorter. In the second model, which includes additional controls for gender and education, the coefficient estimate of subjective remaining life duration remains significant but at a 5 percent level of significance. This model also suggests that, as might be expected, women have a lower mortality risk than men. The *p*-value of the Wald test for education also indicates that the coefficients of high and low education are jointly significant but only at a 10 percent level of significance. In the third model, to which we add standardized household income terciles and marital status as control variables, the coefficient of the subjective remaining life duration is still significant at a 5 percent level of significance. Respondent's income level and marital status have no significant effect on actual mortality.

When health indicators are added to the set of covariates, in contrast, the coefficient of subjective remaining life duration becomes smaller and is no longer statistically significant. In other words, the predictive power of SSPs in explaining mortality disappears once health indicators are controlled for. This finding suggests that SSPs' predictive power for actual mortality, being largely determined by the observed current health situation, is rather limited and contains little additional information on future health expectations. Among the health indicators, smoking, drinking, being in good health, obesity, and having chronic illnesses are statistically significant determinants of mortality risk, with the coefficients of obesity and overweight being jointly significant. The coefficient of year dummy for 1996 is insignificant, but as is to be expected, in all models, mortality risk increases significantly with age.

We then use the coefficient estimates of the third and the fourth models in Table 6 to predict the median remaining life duration conditional on baseline age (see equation (A8), Appendix A). The first panel in Table 7 shows the response of predicted median remaining life duration based on observed mortality to a change in the subjective median remaining life duration of five years. In the model controlling for socioeconomic variables and health indicators, for 45-year-old men reported their SSP in 1995, a five-year change in the subjective median remaining life duration (from 34 to 39 years) results in a change in the predicted median remaining life duration of only 0.22 years.[12] That this difference is statistically insignificant is not surprising given that subjective remaining life duration can no longer predict mortality risk once health indicators are controlled for. When health indicators are excluded, the correlation becomes stronger: for 45-year-old men reported their SSP in 1995, five years of additional subjective remaining life duration corresponds to about one year longer life duration, an increase that is statistically significant.

The second panel in Table 7 reports the male-female difference in predicted remaining median life duration implied by the objective mortality model. It shows a 6.4 years higher predicted median for 45-year-old women than for men of the same age, a significantly positive difference. Using the 1995 HMD life tables, we find that the male-female difference in median life expectancy at age 45 is 6.05 years, only slightly smaller than our model predicts. This proximity suggests some confidence in our model and the estimated age and gender effects.[13]

[12] Predictions are virtually the same for 45-year-old women, which can be expected from the use of a proportional hazard specification.

[13] As already pointed out, if mortality rates decline over time, the period life tables used are likely to underestimate life expectancy compared to cohort life tables.

4.2. Objective and Subjective Mortality Risk Models

This section reports the estimation results for the subjective and objective mortality risk models when we include the same socioeconomic variables and health indicators in each model. If respondents in the sample are able to predict their remaining lifetime correctly, we expect that the signs and magnitudes of the estimates obtained from the objective mortality model will coincide with those obtained from the subjective mortality.

The first two models reported in Table 8 explain subjective and objective mortality risk as a function of year dummy, age effects, and socioeconomic variables. In both models, the coefficient estimate of age has the same sign and is statistically significant although age gradient is steeper in subjective than in actual mortality. The coefficient of female in the subjective model is insignificant and this suggests that males and females have similar beliefs about survival probabilities. The coefficients of income terciles in the subjective model have opposite signs meaning that, taking household size and composition into account, those living in a high income household have significantly higher mortality risks than those living in a middle income household. The estimated coefficient on high education indicates that more educated have higher mortality risks than medium educated although it is statistically insignificant. This finding is at odds with Delavande and Rohwedder (2011) which finds that college graduates report higher survival chances than individuals with lower education. According to Table 8, contrary to objective model's predictions, widowed individuals have lower mortality risks than married individuals. The third and the fourth models show the estimated coefficients on year dummy, gender, age, and health indicators. Among the health indicators, smoking, drinking, being in good health, and obesity explain subjective mortality risk, yet their associations with subjective survival chances are less strong than with objective survival. Specifically, the estimated coefficients on smoking, drinking, and obesity in the subjective model are smaller than those in the objective model, which suggests that individuals underestimate the risk from smoking, obesity and alcohol consumption.

In the last two models in Table 8, which control for both socioeconomic variables and health indicators, the coefficients of age, smoking, good health, and obese have the same sign and are all significant. Even after controlling for socioeconomic variables, the results suggest that individuals underestimate behavioral risks like smoking, obesity, and alcohol consumption when they answer survival probability questions. This finding is in line with the study by Hurd (2009) which shows that, in the U.S., smokers and heavy drinkers overestimate their survival chances relative to their actual survival rates within the sample. In this richer model specification, the coefficients on highest income tercile and widowed still have unexpected signs and they are statistically significant at 5 and 10 percent, respectively. Moreover, the *p*-value of the Wald test indicates that the coefficients of income, education, and marital status are jointly significant at 5 percent in the subjective model.

Next we examine whether respondents over- or underestimate their median remaining lifetime by using the coefficient estimates in columns (5) and (6) of Table 8 (see equation (A9) and (A10), Appendix A) to estimate the predicted median remaining life duration implied by the objective and subjective mortality models. According to the results in Table 9, although both male and female 45-year-olds underestimate their remaining life duration, women tend to do so much more than men. The underestimation for a reference man and woman is 1.2 and 8 years, respectively. Similarly, Teppa (2012) finds that both Dutch men and women have lower SSPs relative to actuarial survival probabilities, on average, while women underestimate their survival chances more than men. Moreover, Perozek (2008) noted that, based on the fitted survival functions, women are pessimistic about their survival chances relative to men in the United States.

Table 9 also shows that, among 45-year-old men, those living in a low (high) income household underestimate their remaining lifetime slightly less (more) than those living in a middle income household, and highly educated men underestimate their remaining lifetime more than medium educated men. Men with a low education level also seem to predict their remaining lifetime better than those with a medium education level, although Table 8 shows that corresponding parameter estimates are not significantly different from zero, hence these differences should be interpreted with some caution.

According to Table 9, both 45-year-old healthy males and unhealthy (reference) males of the same age underestimate their remaining lifetime, yet the prediction error is larger in the case of healthy males. This finding is consistent with the results in Table 8—specifically, a higher (negative) estimated coefficient in the subjective than in the objective model—healthy tend to overestimate their mortality risk more than unhealthy individuals. Table 8 also shows that estimated coefficient on smoking is lower in the subjective than in the objective model, meaning that smokers underestimate their mortality risk. As a result, they significantly overestimate their remaining lifetime compared to non-smokers, as it is shown in Table 9. Similarly, being obese, having chronic illnesses, and consuming alcohol result in overestimation of the remaining lifetime.

5. Conclusions

Our research of whether individuals' subjective survival probabilities (SSPs) convey useful information on their actual (objective) mortality produces several important findings. In our sample, drawn from the DNB Household Survey, Dutch life expectancy as measured by SSPs does indeed predict actual mortality in models that control for income and education level. This predictive power disappears, however, when we control for strong health indicators like self-rated health and smoking behavior. This finding supports the contention that SSPs' predictive power may be limited by their dependence on perceived current health status, meaning that they offer little additional information on expected future health events. Yet, SSPs do correlate with the determinants of objective (actual) mortality. Dutch men and women underestimate, on average, their remaining life duration by about 1 and 8 years, respectively. This underestimation has implications for individuals' long term decisions, such as those on retirement and savings, whose outcomes may be suboptimal. We further show that individuals underestimate the risks from smoking, obesity, and alcohol consumption.

Our survey respondents' underestimation of remaining life duration implies important directions for future research. In particular, we believe that more investigation is needed into why Dutch men and, especially, women underestimate their life expectancies. Could it be, for instance, that the concept of SSP is not clear to many respondents or that some respondents lack basic knowledge on actuarial life duration and behavioral health risks? Such questions need to be answered before we can confidently employ SSPs in economic models (e.g., life cycle models of saving, consumption, and retirement) under the assumption that they convey useful information on individuals' beliefs about mortality risks.

Acknowledgments

We wish to thank CentERdata for making the data available to us, and Rob Alessie, Hans van Kippersluis, Sebastian Kluth, Bertrand Melenberg, Peter van Santen and seminar participants at the MESS workshop (Amsterdam, September 2012), the CeRP conference (Turin, September 2012), the Netspar workshop (Den Haag, October 2012), and the Netspar Pension Day (Utrecht, November 2012) for valuable comments and discussions. Financial support has been provided by the Network for Studies on Pensions, Aging and Retirement (Netspar).

References

Alessie, R., S. Hochguertel, S. and van Soest, A. 2002. "Household portfolios in the Netherlands," in *Household Portfolios,* edited by L. Guiso, M. Haliassos, and T. Jappelli, pp. 341–388. Cambridge, MA: MIT Press.

De Waegenaere, A., Melenberg, B. and Stevens, R. 2010. "Longevity Risk," *De Economist,* 158:151-192.

Delavande, A. and S. Rohwedder. 2011. "Differential mortality in Europe and the U.S.: Estimates based on subjective probabilities of survival," *Demography*, 48: 1377–1400.

Gan, L., Gong, G., Hurd, M. and McFadden, D. 2004. "Subjective mortality risk and bequests," NBER Working Paper No.10789.

Gompertz, B. 1825. "On the nature of the function expressive of the Law of Human Mortality and on a new mode of determining life contingencies," *Philosophical Transactions of the Royal Society of London*, 115: 513–585.

Hamermesh, D. S. 1985, “Expectations, life expectancy, and economic behavior,” *Quarterly Journal of Economics*, 100: 389–408.

Hurd, M. 2009. “Subjective Probabilities in Household Surveys,” *The Annual Review of Economics*, 1: 543–562.

Hurd, M. D. and McGarry. K. 1995. “Evaluating subjective probabilities of survival in the Health and Retirement Study,” *Journal of Human Resources*, 30 (Supplement): S268-S292.

Hurd, M. and McGarry. K. 2002. “The predictive validity of subjective probabilities of survival,” *Economic Journal*, 112: 966–985.

Kleinjans, K.J. and van Soest, A. 2013. “Rounding, focal point answers, and nonresponse to subjective probability questions,” *Journal of Applied Econometrics*, forthcoming.

Lazarsfeld, P.F. 1940. “Panel studies,” *Public Opinion Quarterly*, 4: 122–128.

Manski, C.F. 2004. “Measuring expectations,” *Econometrica*, 72: 1329–1376.

O’Donnell O, Teppa F, and Van Doorslaer E. 2008. “Can subjective survival expectations explain retirement behavior?” DNB Working Paper No. 188.

Olshansky, S.J. and Carnes, A.C. 1997. “Ever since Gompertz,” *Demography*, 34(1): 1-15.

Payne, W.J, Sagara, N., Shu, S.B., Appelt, K.C., and Johnson, E.J. 2013 “Life expectancy as a constructed belief: Evidence of a live-to or die-by framing effect,” *Journal of Risk and Uncertainty*, 46(1): 27–50.

Peracchi, F. and Perotti, V., 2011. “Subjective survival probabilities and life tables: Evidence from Europe,” EIEF Working Paper No.1016.

Perozek, M. 2008. “Using subjective expectations to forecast longevity: Do survey respondents know something we don’t know?” *Demography*, 45: 95–113.

Salm, M. 2010. “Subjective mortality expectations and consumption and savings behaviours among the elderly,” *Canadian Journal of Economics*, 43(3): 1040–1057.

Siegel, M., Bradley, E.H., and Kasl, S.V. 2003. “Self-rated life expectancy as a predictor of mortality: Evidence from the HRS and AHEAD surveys,” *Gerontology* 49: 265–71.

Siermann, C., van Teeffelen, P., and L. Urlings. 2004. “Equivalentiefactoren 1995-2000: Methode en belangrijkste uitkomsten,” *Sociaal-economische trends*, 3: 63–66.

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

Sturgis, P., Allum, N. and Brunton-Smith, I. 2009. “Attitudes over time: The psychology of panel conditioning,” in *Methodology of Longitudinal Surveys*, edited by P. Lynn, pp. 113–126. Chichester: Wiley.

Teppa, F. 2012. “Can the longevity risk alleviate the annuitization puzzle? Empirical evidence from Dutch data,” DNB Working Paper No.302.

Van Doorn, C. and Kasl, S. V. 1998. “Can parental longevity and self-rated life expectancy predict mortality among older persons? Results from an Australian cohort,” *Journal of Gerontology: Social Sciences*, 53B(1): S28–S34.

Zellner, A. 1962. “An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias,” *Journal of the American Statistical Association,* 57: 348-368.

Figures and Tables

Figure 1: Average subjective remaining life duration across age categories (in years).

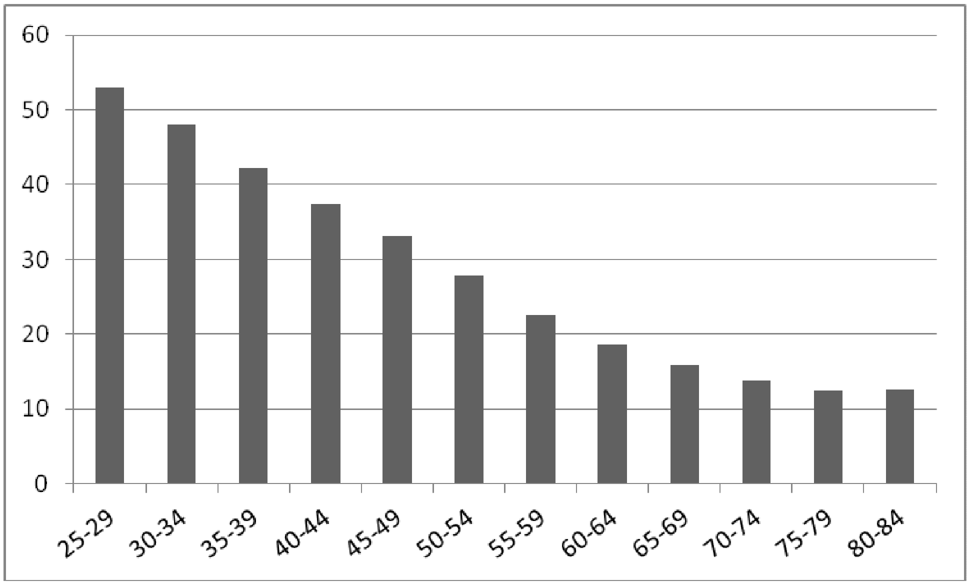

Table 1: Response Rates to Survival Probabilities

Age of respondent	Survival probabilities	Number of respondents asked the probability questions	Number of respondents who actually answered the probability questions	Response rate (%)
16-65	P75 and P80	5,130	4,427	86.296
65-70	P80 and P85	307	267	86.971
70-75	P85 and P90	186	158	84.946
75-80	P90 and P95	83	74	89.157
80-85	P95 and P100	32	27	84.375

Notes: N = 5,738. Nine observations were excluded from the analysis for being in the 85+ age category, for which we do not observe the two survival probabilities.

Table 2: Inconsistency Rates to Survival Probabilities

Probability comparison	Number of respondents	% of respondents
Prob($t_{1,i}$)<Prob($t_{2,i}$)	31	0.626
Prob($t_{1,i}$)=Prob($t_{2,i}$)	1,590	32.102
Prob($t_{1,i}$)>Prob($t_{2,i}$)	3,332	67.272

Notes: $t_{1,i}$ and $t_{2,i}$ are two different target ages, where $t_{1,i} < t_{2,i}$; 4,953 individuals out of 5,747 reported both Prob($t_{1,i}$) and Prob($t_{2,i}$).

Table 3: The Mean of Objective and Subjective Remaining Life Duration across Age Categories and Gender (in years)

Age	Remaining life duration: Objective* (1)	Remaining life duration: Subjective (2)	(2)-(1)	Remaining life duration: Objective* (3)	Remaining life duration: Subjective (4)	(4)-(3)
	Males			Females		
25-29	53.11	52.71	-0.4	59.84	53.27	-6.57
30-34	48.54	47.66	-0.88	55.49	48.31	-7.18
35-39	43.55	42.34	-1.21	50.25	42.13	-8.12
40-44	38.83	37.3	-1.53	45.49	37.61	-7.88
45-49	33.8	32.83	-0.97	40.47	33.49	-6.98
50-54	29.21	27.44	-1.77	35.72	28.13	-7.59
55-59	24.11	22.42	-1.69	30.58	22.55	-8.03
60-64	19.75	18.47	-1.28	25.92	19.05	-6.87
65-69	15.66	15.82	0.16	21.44	15.79	-5.65
70-74	11.73	13.98	2.25	17.25	13.37	-3.88
75-79	8.71	12.21	3.5	13.54	12.66	-0.88
80-84	6.15	11.93	5.78	10.35	13.3	2.95
Sample average	33.36	32.46	-0.9	41.84	34.68	-7.16

Notes: N = 3,059. * The objective remaining life duration is computed based on the estimation of a Gompertz mortality model with only gender and age as covariates.

Table 4: The Mean of the Objective and Subjective Remaining Life Duration across Education, Health Status, and Gender (in years)

	Remaining life duration			Remaining life duration		
Age	Objective (1)	Subjective (2)	(2)-(1)	Objective (3)	Subjective (4)	(4)-(3)
	Males			Females		
Low education	31.13	30.99	-0.14	39.46	32.19	-7.27
Medium education	35.30	33.20	-2.10	43.72	35.36	-8.36
High education	35.73	32.39	-3.34	46.50	36.29	-10.21
In good health*	37.14	33.97	-3.17	46.50	36.26	-10.24
Not in good health*	25.66	24.92	-0.74	35.21	28.55	-6.66

Notes: N = 3,059. *Both education and health are controlled for to obtain the objective remaining life duration.

Table 5: Mean, Median, and Standard Deviation of Variables

Variable	Mean	Median	SD
Gender (0 = male,1 =female)	0.450	0	0.498
Year dummy 1996	0.161	0	0.367
Subjective remaining life duration (in years)	33.460	34.245	12.598
Smoking	0.312	0	0.463
Good health	0.816	1	0.387
Alcohol	0.081	0	0.272
Chronic illness	0.245	0	0.430
Overweight	0.336	0	0.473
Obese	0.060	0	0.237
Low education	0.225	0	0.418
High education	0.400	0	0.490
Standardized income (in Dutch guilders)	44,287	40,757	24,988
Married	0.876	1	0.330
Single	0.100	0	0.300
Widowed	0.020	0	0.154
Age (in years)	47.321	45.833	12.654

Notes: N = 3,059.

Table 6: Estimation Results for the Objective Mortality Model

Covariates	(1)	(2)	(3)	(4)
Subjective remaining life duration	-0.002***	-0.002**	-0.002**	-0.0004
	(0.001)	(0.001)	(0.001)	(0.0008)
Year dummy 1996	0.041	0.034	0.035	0.063
	(0.187)	(0.188)	(0.188)	(0.189)
Female		-0.777***	-0.773***	-0.778***
		(0.127)	(0.130)	(0.133)
Low education		0.100	0.109	0.163
		(0.140)	(0.141)	(0.141)
High education		-0.207	-0.130	-0.040
		(0.130)	(0.137)	(0.139)
Lowest standardized income tercile			0.116	0.038
			(0.133)	(0.135)
Highest standardized income tercile			-0.107	-0.043
			(0.142)	(0.143)
Single			0.221	0.097
			(0.186)	(0.188)
Widowed			0.136	0.103
			(0.206)	(0.207)
Smoking				0.728***
				(0.124)
Good health				-0.443***
				(0.138)
Alcohol				0.297*
				(0.173)
Overweight				-0.089
				(0.120)
Obese				0.465**
				(0.191)
Chronic illness				0.260*
				(0.133)
Constant	-10.531***	-10.396***	-10.415***	-11.471***
	(0.520)	(0.523)	(0.527)	(0.561)
Age	0.008***	0.008***	0.008***	0.009***
	(0.001)	(0.001)	(0.001)	(0.001)
Log likelihood	-351.264	-330.569	-328.477	-292.933
p-value Wald test: education		0.092	0.302	0.377
p-value Wald test: education and income			0.141	0.605

p-value Wald test: education, income, marital status	0.169	0.787
p-value Wald test: overweight and obese		0.019

Notes: N = 3,059; no. of failures = 330; standard errors given in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 7: Predicted Median Remaining Life Duration based on Estimates from the Objective Mortality Model (in years)

	Baseline age	Subjective remaining life duration		Predicted objective remaining life duration		Change in predicted remaining life duration	
		(a)	(b)	(a)	(b)	Point estimate (b)-(a)	p-value*
Panel A: Change in male subjective remaining life duration							
With health controls [a]	45	34	39	38.872	39.092	0.220	0.605
No health controls [b]	45	34	39	40.784	41.793	1.009	0.054
Panel B: Gender differences without control variables							
				Male (a)	Female (b)	Point estimate (b)-(a)	p-value*
	45	-	-	37.676	44.085	6.409	0.000

Notes: N = 3,059. [a]The reference category is a married man living in a medium income household, who reported his SSP in 1995,with no chronic illnesses, non-smoker, non-drinker, of normal weight, medium-level education, but not in good health. [b]The reference category is a married, medium educated man living in middle income household, who reported his SSP in 1995.*p-values are the result of a two-tailed t-test.

Table 8: Estimation Results for the Objective and Subjective Mortality Risk Models

	Subjective	Objective	Subjective	Objective	Subjective	Objective
Covariates	(1)	(2)	(3)	(4)	(5)	(6)
Year dummy 1996	0.026	0.391	0.001	0.086	0.024	0.065
	(0.056)	(0.188)	(0.053)	(0.188)	(0.059)	(0.188)
Female	0.005	-0.788***	0.001	-0.735***	0.012	-0.782***
	(0.042)	(0.129)	(0.040)	(0.127)	(0.045)	(0.133)
Low education	0.007	0.111			-0.010	0.164
	(0.054)	(0.141)			(0.057)	(0.141)
High education	0.027	-0.142			0.060	-0.041
	(0.048)	(0.137)			(0.052)	(0.139)
Lowest standardized income tercile	-0.018	0.098			-0.021	0.034
	(0.050)	(0.133)			(0.053)	(0.135)
Highest standardized income tercile	0.094*	-0.119			0.118**	-0.042
	(0.050)	(0.142)			(0.053)	(0.143)
Single	-0.024	0.236			-0.065	0.098
	(0.068)	(0.186)			(0.072)	(0.188)
Widowed	-0.239*	0.142			-0.276*	0.106
	(0.132)	(0.206)			(0.141)	(0.206)
Smoking			0.162***	0.741***	0.179***	0.730***
			(0.043)	(0.123)	(0.047)	(0.124)
Good health			-0.220***	-0.465***	-0.239***	-0.455***
			(0.058)	(0.135)	(0.064)	(0.136)
Alcohol			0.130*	0.305*	0.124	0.302*
			(0.074)	(0.171)	(0.081)	(0.173)
Overweight			-0.016	-0.074	-0.007	-0.089
			(0.043)	(0.119)	(0.047)	(0.120)
Obese			0.175**	0.493***	0.122**	0.470**
			(0.084)	(0.189)	(0.093)	(0.191)
Chronic illness			0.073	0.281**	0.082	0.266**
			(0.053)	(0.131)	(0.057)	(0.132)
Constant	-14.543***	-11.409***	-14.411***	-11.730***	-14.456***	-11.711***
	(0.173)	(0.272)	(0.178)	(0.308)	(0.186)	(0.335)
Age	0.014***	0.009***	0.014***	0.010***	0.014***	0.010***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)
p-value Wald test:	0.069	0.356	-	-	0.027	0.886

income						
p-value Wald test: education, income, and marital status	0.104	0.142	-	-	0.010	0.789
p-value Wald test: BMI			0.086	0.014	0.059	0.017

Notes: N = 3,059; no. of failures = 330; standard errors given in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 9: Objective versus Subjective Predicted Life Duration (in years)

Characteristics	Objective predicted remaining life duration (1)	Subjective predicted remaining life duration (2)	(2)-(1)
Predictions using SES and HEALTH variables			
Man (reference) [a]	38.113	36.869	-1.244
Woman	44.829	36.801	-8.028
Man living in a low income household	37.826	36.990	-0.836
Man living in a middle income household	38.113	36.869	-1.244
Man living in a high income household	38.474	36.190	-2.284
Low educated man	36.713	36.925	0.212
Medium educated man	38.113	36.869	-1.244
High educated man	38.467	36.523	-1.944
Smoking man	31.907	35.838	3.931
Man in good health	42.013	38.249	-3.765
Man drinking alcohol	35.538	36.154	0.617
Man with chronic illnesses	35.839	36.398	0.559
Obese man	34.112	35.636	1.524

[a]The reference is a 45-year old married man living in a middle income household who reported his SSP in 1995, with no chronic illnesses, non-smoker, non-drinker, of normal weight and medium education, but not in good health.

Appendix A

A1. Derivation of the Median Remaining Life Duration

For each individual in our sample, we observe two values of the survival function, $SSP_{1,i}$ and $SSP_{2,i}$, at two different target ages $t_{1,i}$ and $t_{2,i}$ where $t_{1,i} < t_{2,i}$. $t_{0,i}$ is the baseline age; that is, the age at which the respondent reports the SSPs:[14]

$$S_{1,i}(t_{0,i}, t_{1,i}) = \exp\left\{ -\int_{t_{0,i}}^{t_{1,i}} \theta(s)ds \right\} = SSP_{1,i} \qquad \text{(A1)}$$

$$S_{2,i}(t_{0,i}, t_{2,i}) = \exp\left\{ -\int_{t_{0,i}}^{t_{2,i}} \theta(s)ds \right\} = SSP_{2,i} \qquad \text{(A2)}$$

with $(t_{1,i}, t_{2,i}) \in \{(75,80),(80,85),(85,90),(90,95),(95,100)\}$ and $(SSP_{1,i}, SSP_{2,i}) \in \{(P75, P80),(P80, P85),(P85, P90),(P90, P95),(P95, P100)\}$, respectively. $P75$ represents the subjective survival probability (SSP) to age 75, $P80$ that to age 80, and so on.

[14] In the mortality hazard model, we measure time in months. Because the initial age in our sample is 25, we are interested in survival from $t_{0,i} - 25$ to $t_i - 25$ where t_i is respondent age at death (if deceased) or respondent age at the end of the DO survey (if still alive at the end of the observation period). Target ages are changed accordingly. For example, instead of $t_{1,i}$ =75, we use $t_{1,i}$ =75*12-25*12=600. Similarly, $t_{0,i} = t_{0,i}$ in months-300. For the sample (50+), instead of $t_{1,i}$ =75 we use $t_{1,i}$ =75*12-50*12=300, and $t_{0,i} = t_{0,i}$ in months-600.

The Gompertz hazard function is given by

$\theta(t) = \lambda_i \exp\{\gamma_i t\}$, $\gamma_i > 0$ for each i, meaning that hazard increases over time (positive duration dependence).

After substituting the hazard rate into equations (A1) and (A2), we evaluate the integral to find

$$S_{1,i}\left(\gamma_i, \lambda_i \middle| t_{0,i}, t_{1,i}\right) = \exp\left\{\frac{\lambda_i}{\gamma_i}\left(\exp\{\gamma_i t_{0,i}\} - \exp\{\gamma_i t_{1,i}\}\right)\right\} = SSP_{1,i} \tag{A3}$$

$$S_{2,i}\left(\gamma_i, \lambda_i \middle| t_{0,i}, t_{2,i}\right) = \exp\left\{\frac{\lambda_i}{\gamma_i}\left(\exp\{\gamma_i t_{0,i}\} - \exp\{\gamma_i t_{2,i}\}\right)\right\} = SSP_{2,i} \tag{A4}$$

Following Perozek (2008), we take logarithms of the survival functions (equation (A5)) and estimate the parameters γ_i and λ_i for each individual using nonlinear least squares (NLLS). This procedure requires that we replace survival probabilities of 0 and 1 with slightly different numbers, namely, 0.01 and 0.99, respectively.

$$\ln\left(SSP_{j,i}\right) = \ln\left(S_{j,i}\left(\gamma_i, \lambda_i \middle| t_{0,i}, t_{j,i}\right)\right) + \varepsilon_{j,i} \qquad j \in \{1,2\} \tag{A5}$$

where $\varepsilon_{j,i}$ is the error term, which is assumed to be independent, identically distributed (i.i.d.), and homoscedastic, and have a zero mean.

The NLLS estimates of γ_i^* and λ_i^* are obtained by minimizing the following expression:

$$\min_{\gamma_i, \lambda_i} \sum_j \left(\ln\left(SSP_{j,i}\right) - \ln\left(S_{j,i}\left(\gamma_i, \lambda_i \middle| t_{0,i}, t_{j,i}\right)\right)\right)^2$$

with $S_{j,i}\left(\gamma_i, \lambda_i \middle| t_{0,i}, t_{j,i}\right) = \exp\left\{\frac{\lambda_i}{\gamma_i}\left(\exp\{\gamma_i t_{0,i}\} - \exp\{\gamma_i t_{j,i}\}\right)\right\}$

Next, we calculate the median remaining life duration conditional on baseline age for each individual (RL_i^S) based on the following formula:[15]

$$S(RL_i^S \mid t_{0,i}) = \exp\left\{-\int_0^{RL_i^S} \theta(s' + t_{0,i})ds'\right\} = 0.5 \tag{A6}$$

The Gompertz hazard function is $\theta(s' + t_{0,i}) = \lambda_i \exp\{\gamma_i (s' + t_{0,i})\}$, $\gamma_i > 0$ for each i.

Evaluating the integral in equation (A6) and taking the natural logarithm of both sides yields

$$-\ln(0.5) = \frac{\lambda_i}{\gamma_i}\left(\exp\{\gamma_i (t_{0,i} + RL_i^S)\} - \exp\{\gamma_i t_{0,i}\}\right)$$

$$RL_i^S = \frac{1}{\gamma_i}\ln\left(\frac{\gamma_i \ln(2)}{\lambda_i \exp(\gamma_i t_{0,i})} + 1\right) \tag{A7}$$

In equation (A7), we replace γ_i and λ_i with their estimates γ_i^* and λ_i^*, respectively. Because the variable, RL_i^S , is created using individuals' subjective survival probabilities, it represents the subjective median remaining life duration conditional on baseline age for each individual.

Predictions Using the Estimates of the Objective Mortality Model (Table 7)

The assumption maintained in the objective mortality model is that there is a Gompertz hazard function given by

$\theta(t \mid \mathbf{x_i}) = \lambda_i \exp\{\gamma_i t\}$, where $\lambda_i = \lambda_O = \exp\{\mathbf{x_i \beta_O}\}$ and $\gamma_i = \gamma_O > 0$

By combining this assumption with equation (A7) and replacing λ_O and γ_O with the estimates of the objective mortality model given in Table 6, we can derive the following equation:

[15] We calculate the median remaining life duration conditional on baseline age because respondents report their SSP knowing that they have survived up to their current age.

$$\hat{RL}^{O} = \frac{1}{\hat{\gamma}_O} \ln\left(\frac{\hat{\gamma}_O \ln(2)}{\exp\left\{ \bar{\mathbf{x}}_\mathbf{i} \hat{\boldsymbol{\beta}}_\mathbf{O} + \hat{\gamma}_O \bar{t}_{0,i} \right\}} + 1 \right) \tag{A8}$$

where $\hat{RL}^{O}$ is the objective predicted remaining life duration and $\bar{t}_{0,i}$ is the baseline age, which is equal to 45. The vector $\bar{\mathbf{x}}_\mathbf{i}$ contains the mean value of the logarithm of household income and fixed values of the remaining variables shown in Table 7. For example, in the first row of Table 7, all of the dummy variables in the objective mortality model are equal to zero, subjective remaining life duration takes a value of 410 or 470, and birth year is equal to 1950.

Comparison of the Objective and the Subjective Predicted Life Durations (Table 9)

The common assumption in the objective and subjective mortality models is that life duration can be modeled using a Gompertz distribution. Under this assumption, the hazard function can be written as

$$\theta(t|\mathbf{x}_\mathbf{i}) = \lambda_i \exp\{\gamma_i t\}$$

where $\lambda_i = \lambda_O = \exp\{\mathbf{x}_\mathbf{i}\boldsymbol{\beta}_\mathbf{O}\}$, $\gamma_i = \gamma_O > 0$ in the objective mortality model, and $\lambda_i = \lambda_S = \exp\{\mathbf{x}_\mathbf{i}\boldsymbol{\beta}_S\}$, $\gamma_i = \gamma_S > 0$ in the subjective mortality model. In equation (A7), we replace λ_O, γ_O, λ_S, and γ_S with their estimates, $\hat{\lambda}_O$, $\hat{\gamma}_O$, $\hat{\lambda}_S$, and $\hat{\gamma}_S$, respectively:

$$\hat{RL}^{O} = \frac{1}{\hat{\gamma}_O} \ln\left(\frac{\hat{\gamma}_O \ln(2)}{\exp\left\{ \bar{\mathbf{x}}_\mathbf{i} \hat{\boldsymbol{\beta}}_\mathbf{O} + \hat{\gamma}_O \bar{t}_{0,i} \right\}} + 1 \right) \tag{A9}$$

$$\hat{RL}^{S} = \frac{1}{\hat{\gamma}_S} \ln\left(\frac{\hat{\gamma}_S \ln(2)}{\exp\left\{ \bar{\mathbf{x}}_{\mathbf{i}} \hat{\boldsymbol{\beta}}_{\mathbf{S}} + \hat{\gamma}_S \bar{t}_{0,i} \right\}} + 1 \right) \tag{A10}$$

where $\hat{RL}^{O}$ and $\hat{RL}^{S}$ denote objective and subjective predicted remaining life durations, respectively, and $\bar{t}_{0,i}$ is the baseline age, which is equal to 45. The vector $\bar{\mathbf{x}}_{\mathbf{i}}$ contains the mean value of the logarithm of household income and fixed values of the remaining variables shown in Table 9. For example, in the first row of Table 9, all of the dummy variables in both mortality models are equal to zero and birth year is equal to 1950.

Appendix B

Table B1: Tabulation of Survival Probabilities (%)

	0.01	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.99
P75	1.67	0.58	1.64	1.73	2.95	27.44	11.56	16.06	22.6	6.72	7.03
P80	4.27	2.14	5.41	8.63	10.74	30.09	16.7	13.39	6.72	1.68	X
P85	6.32	3.57	8.24	13.74	11.81	31.87	10.16	6.59	4.67	X	X
P90	19.23	11.54	13.94	10.58	9.13	25.96	X	X	X	X	X

Notes: N = 3,059. X means that the number in this cell cannot be released by the CBS since the number of units underlying this cell is less than 10. We do not report survival probabilities to age 95 or age 100 because the number of observations is so small. For the reasons given in Appendix A, we replace survival probability answers 0 and 1 with 0.01 and 0.99, respectively.

Table B2: Variable Definitions

Variable	Description	Baseline
RL_i^S	Subjective median remaining life duration conditional on baseline age	---
Female	Respondent is female	Respondent is male
Low education	Respondent has primary/low level education or vocational training through the apprentice system	Respondent has pre-university education or junior/senior vocational training

High education	Respondent has a university degree or a vocational college degree.	Respondent has pre-university education or junior/senior vocational training
Standardized household income (in Dutch guilders)	The sum of the net annual incomes of all household members after deduction of taxes but before making payments such as rent, mortgages, etc. divided by the equivalence scale provided by Statistics Netherlands (Siermann *et al*., 2004)	---
Lowest standardized income tercile	Respondent's standardized household income is lower than and equal to the 33th percentile of the standardized household income series	Respondent's standardized household income is between the 33th and 67th percentiles of standardized household income series
Highest standardized income tercile	Respondent's standardized household income is higher than and equal to the 67th percentile of the standardized household income series	Respondent's standardized household income is between the 33th and 67th percentiles of standardized household income series
Good health	Respondent's self-reported health is excellent/good	Respondent's self-reported health is fair/not so good/poor
Smoking	Respondent is smoker	Respondent is non-smoker
Alcohol	Respondent consumes more than 4 alcoholic drinks a day	Respondent does not consume more than 4 alcoholic drinks a day
Chronic illness	Respondent suffers from long- term illness, disorder, disability, or the consequences of an accident	Respondent does not suffer from long-term illness, disorder, disability, or the consequences of an accident
Overweight	25≤ Respondent's body mass index (BMI)< 30	Respondent's BMI< 25
Obese	Respondent's BMI≥ 30	Respondent's BMI< 25
Year dummy 1996	Respondent reported his/her SSP in 1996	Respondent reported his/her SSP in 1995
Single	Respondent is single	Respondent is married
Widowed	Respondent is widowed	Respondent is married
Age	Respondent's age in months at the time of interview	---

Appendix C

Estimation results are based on the sample aged 25+ with respondents reporting equal survival probabilities included. We assume a 5 percent shift between equal survival probabilities.

Table C1: Estimation Results for the Objective Mortality Model

Covariates	(1)	(2)	(3)	(4)
Subjective remaining life duration	-0.002***	-0.002***	-0.002***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Year dummy 1996	0.046	0.041	0.043	0.101
	(0.149)	(0.149)	(0.149)	(0.151)
Female		-0.782***	-0.784***	-0.766***
		(0.105)	(0.107)	(0.109)
Low education		0.065	0.069	0.065
		(0.115)	(0.116)	(0.117)
High education		-0.281**	-0.234**	-0.152
		(0.111)	(0.118)	(0.119)
Lowest standardized income tercile			0.174	0.104
			(0.118)	(0.119)
Highest standardized income tercile			0.036	0.050
			(0.119)	(0.119)
Single			0.249	0.107
			(0.155)	(0.157)
Widowed			0.107	0.052
			(0.177)	(0.177)
Smoking				0.662***
				(0.104)
Good health				-0.374***
				(0.118)
Alcohol				0.283*
				(0.146)
Overweight				-0.004
				(0.100)
Obese				0.428**
				(0.167)
Chronic illness				0.239**
				(0.112)
Constant	-10.048***	-9.789***	-9.863***	-10.656***
	(0.362)	(0.363)	(0.369)	(0.389)

Age	0.007***	0.008***	0.007***	0.008***
	(0.000)	(0.000)	(0.000)	(0.000)
Log likelihood	-538.612	-507.866	-505.239	-464.522
p-value Wald test: education		0.009	0.050	0.244
p-value Wald test: education and income			0.020	0.336
p-value Wald test: education, income, marital status			0.021	0.527
p-value Wald test: overweight and obese				0.029

Notes: N = 4,434; no. of failures = 463; standard errors given in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table C2: Estimation Results for the Objective and the Subjective Mortality Risk Models

	Subjective	Objective	Subjective	Objective	Subjective	Objective
Covariates	(1)	(2)	(3)	(4)	(5)	(6)
Year dummy 1996	0.015	0.063	-0.004	0.131	0.016	0.114
	(0.048)	(0.149)	(0.042)	(0.150)	(0.048)	(0.151)
Female	-0.072*	-0.798***	-0.087***	-0.737***	-0.060	-0.780***
	(0.037)	(0.107)	(0.032)	(0.105)	(0.038)	(0.109)
Low education	0.019	0.058			-0.005	0.054
	(0.047)	(0.116)			(0.047)	(0.116)
High education	0.115***	-0.242**			0.147***	-0.153
	(0.043)	(0.118)			(0.043)	(0.119)
Lowest standardized income tercile	-0.054 (0.045)	0.144 (0.118)			-0.059 (0.045)	0.079 (0.119)
Highest standardized income tercile	0.023	0.024			0.038	0.050
	(0.045)	(0.119)			(0.044)	(0.119)
Single	0.012	0.280*			-0.029	0.112
	(0.061)	(0.155)			(0.061)	(0.157)
Widowed	-0.061	0.087			-0.118	0.051
	(0.118)	(0.177)			(0.118)	(0.177)
Smoking			0.098***	0.689***	0.121***	0.663***
			(0.035)	(0.102)	(0.039)	(0.103)
Good health			-0.397***	-0.450***	-0.411***	-0.438***
			(0.048)	(0.116)	(0.054)	(0.116)
Alcohol			0.162***	0.316**	0.152**	0.321**
			(0.060)	(0.144)	(0.068)	(0.145)
Overweight			0.005	0.018	0.022	0.001
			(0.034)	(0.099)	(0.039)	(0.100)

Obese			0.125*	0.471***	0.171**	0.444***
			(0.068)	(0.166)	(0.078)	(0.167)
Chronic illness			0.063	0.278**	0.072	0.262**
			(0.042)	(0.111)	(0.048)	(0.112)
Constant	-12.834***	-11.227***	-12.525***	-11.495***	-12.589***	-11.451***
	(0.133)	(0.219)	(0.138)	(0.254)	(0.146)	(0.271)
Age	0.011***	0.009***	0.011***	0.009***	0.011***	0.009***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
p-value Wald test: income	0.229	0.431	-	-	0.106	0.793
p-value Wald test: education, income, and marital status	0.019	0.023	-	-	0.000	0.618
p-value Wald test: BMI			0.186	0.015	0.093	0.023

Notes: N = 4,434; no. of failures = 463; standard errors given in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Appendix D

The estimation results are based on the sample aged 50+ with respondents reporting equal survival probabilities excluded.

Table D1: Estimation Results for the Objective Mortality Model

Covariates	(1)	(2)	(3)	(4)
Subjective remaining life duration	-0.002**	-0.002**	-0.002*	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Year dummy 1996	0.062	0.065	0.048	0.094
	(0.214)	(0.215)	(0.215)	(0.217)
Female		-0.837***	-0.830***	-0.832***
		(0.141)	(0.144)	(0.148)
Low education		0.168	0.173	0.234
		(0.153)	(0.154)	(0.154)
High education		-0.124	-0.016	0.072
		(0.142)	(0.151)	(0.154)
Lowest standardized income tercile			-0.203	0.100
			(0.171)	(0.144)
Highest standardized income tercile			0.143	-0.081
			(0.141)	(0.173)
Single			0.064	-0.029

			(0.216)	(0.220)
Widowed			0.096	0.081
			(0.210)	(0.211)
Smoking				0.609***
				(0.142)
Good health				-0.414***
				(0.149)
Alcohol				0.253
				(0.200)
Overweight				-0.021
				(0.131)
Obese				0.463**
				(0.211)
Chronic illness				0.271*
				(0.143)
Constant	-8.371***	-8.238***	-8.262***	-8.915***
	(0.371)	(0.376)	(0.386)	(0.426)
Age	0.009***	0.009***	0.009***	0.010***
	(0.001)	(0.001)	(0.001)	(0.001)
Log likelihood	-354.957	-335.432	-333.191	-310.639
p-value Wald test: education		0.188	0.446	0.313
p-value Wald test: education and income			0.121	0.434
p-value Wald test: education, income, marital status			0.263	0.687
p-value Wald test: overweight and obese				0.068

Notes: $N = 1{,}142$; no. of failures = 277; standard errors given in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table D2: Estimation Results for the Objective and the Subjective Mortality Risk Models

	Subjective	Objective	Subjective	Objective	Subjective	Objective
Covariates	(1)	(2)	(3)	(4)	(5)	(6)
Year dummy 1996	0.011	0.064	0.006	0.143	0.020	0.102
	(0.115)	(0.215)	(0.118)	(0.214)	(0.123)	(0.216)
Female	-0.022	-0.843***	-0.037	-0.802***	-0.026	-0.838***
	(0.075)	(0.144)	(0.076)	(0.140)	(0.082)	(0.147)
Low education	0.025	0.172			0.015	0.234
	(0.092)	(0.153)			(0.099)	(0.154)
High education	0.022	-0.022			0.046	0.072
	(0.088)	(0.151)			(0.095)	(0.154)
Lowest standardized	-0.127	0.138			-0.144	0.098

income tercile	(0.089)	(0.142)			(0.095)	(0.144)
Highest standardized income tercile	0.051	-0.229			0.117	-0.085
	(0.088)	(0.171)			(0.095)	(0.173)
Single	0.101	0.082			0.048	-0.024
	(0.129)	(0.216)			(0.138)	(0.220)
Widowed	-0.171	0.107			-0.152	0.088
	(0.152)	(0.210)			(0.163)	(0.211)
Smoking			0.125	0.618^{***}	0.139	0.611^{***}
			(0.088)	(0.140)	(0.092)	(0.142)
Good health			-0.258^{***}	-0.442^{***}	-0.281^{***}	-0.435^{***}
			(0.096)	(0.145)	(0.101)	(0.147)
Alcohol			0.223^{*}	0.249	0.211	0.261
			(0.132)	(0.197)	(0.138)	(0.199)
Overweight			-0.064	-0.023	-0.055	-0.027
			(0.077)	(0.130)	(0.081)	(0.131)
Obese			0.136	0.482^{**}	0.183	0.466^{**}
			(0.145)	(0.209)	(0.153)	(0.211)
Chronic illness			0.127	0.295^{**}	0.156^{*}	0.277^{*}
			(0.088)	(0.140)	(0.092)	(0.143)
Constant	-10.274^{***}	-8.859^{***}	-10.170^{***}	-9.082^{***}	-10.187^{***}	-9.133^{***}
	(0.178)	(0.242)	(0.193)	(0.277)	(0.210)	(0.312)
Age	0.015^{***}	0.010^{***}	0.014^{***}	0.010^{***}	0.015^{***}	0.010^{***}
	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)
p-value Wald test: income	0.140	0.109	-	-	0.035	0.576
p-value Wald test: education, income, and marital status	0.335	0.201	-	-	0.128	0.685
p-value Wald test: BMI			0.361	0.051	0.292	0.063

Notes: $N = 1,142$; no. of failures = 277; standard errors given in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.