Network for Studies on Pensions, Aging and Retirement

NETSPAR ACADEMIC SERIES

Does respondent's knowledge on population life expectancy influence the accuracy of subjective survival probabilities?

Vesile Kutlu-Koc and Adriaan Kalwij

DP 02/2017-004

Does respondent's knowledge on population life expectancy influence the accuracy of subjective survival probabilities?

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

February 27, 2017

Abstract

Life expectancy plays a role in many decisions individuals take. Previous studies have shown that, on average, individuals underestimate their remaining life expectancy and this could yield suboptimal outcomes. Using a Dutch Household Survey supplemented with administrative data on mortality we find that individuals predict significantly more accurate their remaining life expectancy when they have better knowledge on population remaining life expectancy. This suggests that informing individuals about population remaining life expectancies may help them making better long term decisions.

Keywords: mortality risk, subjective survival, survey data, sample selection model

[a] Corresponding author. kutlukoc@mea.mpisoc.mpg.de, Munich Center for the Economics of Aging (MEA), Amalienstr. 33, Room. 360, 80799 Munich, Germany.

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

1. Introduction

Individuals' decisions such as when to retire, whether to purchase a life insurance policy, when to move to an old age home or whether to adopt a healthy lifestyle, may be influenced by their beliefs on remaining lifetime. Theoretical economic models of life cycle behavior such as Hurd (1989) have shown the importance of allowing for lifetime uncertainty for the savings behavior of elderly singles and empirical support for his model has been presented in Gan *et al*. (2004), Kutlu-Koc et al. (2014) and Salm (2010). In household surveys, individual life expectancy is usually elicited using questions about the probability of survival to a target age (Manski 2004). Based on such subjective survival probabilities (SSPs) several studies have shown that individuals on average underestimate their life expectancy (see, e.g., Perozek 2008; Teppa 2012; Groneck *et al.* 2013). For the Netherlands, the country also analyzed in this study, Kutlu-Koc and Kalwij (2017) found that, on average, women underestimate their remaining life expectancy, whereas men tend to estimate it more accurately. Under- or overestimation of remaining life expectancy by individuals may yield suboptimal outcomes. For example, those who think that they will not live until old age may not save enough to finance consumption during retirement or may move to an old age home too soon.

An important public policy issue we address in this paper is if individuals' under- or overestimation of remaining life expectancy is related to insufficient knowledge on population life expectancy. If, for instance, we would find that people who underestimate population remaining life expectancy also underestimate their remaining life expectancy, this may suggest that policymakers could improve individuals' decisions by informing them on population remaining life expectancy based on age and gender.

Two previous studies suggest that respondents' knowledge on population life expectancy could affect their beliefs on remaining lifetime. Elder (2013) found that US respondents made more precise subjective survival forecasts, i.e. uncertainty in their forecasts decreased, after having received information about of population survival rates. Steffen (2009) combined the elicited beliefs of German respondents on population life expectancy and their own position relative to it. About two-thirds of the respondents expected to live about as long as the self-reported population life expectancy. However, their self-reported population life expectancy was, on average, significantly lower than actuarial life expectancy. These findings of Steffen (2009) suggest that individuals' underestimation (on average) of their own remaining life expectancy could be related to their underestimation of population life expectancy. Both studies, however, do not provide evidence on the relationship between people's knowledge about population life expectancy and the closeness of their subjective to their objective survival rates.[1]

Our main contribution to the empirical literature is that we show that people who have better knowledge on population life expectancy predict significantly more accurate their remaining life expectancy in the sense that these predictions are, on average, closer to their objective remaining life expectancies. For our analysis, we use Dutch data from the 1995 and 1996 DNB Household Survey (DHS) supplemented with administrative data on (objective) mortality from the causes of death registry over the period 1995-2010. That is, from administrative records we observe if individuals have died before the end of 2010. We calculate subjective life expectancies based on two subjective probabilities of survival for each individual

[1] A conceptually related study of McFadden *et al.* (2004) shows that US respondents reported a lower subjective probability (and, on average, closer to the objective probability) of being admitted to a nursing home when they had received information about the fraction of people admitted to nursing homes at the beginning of the survey.

in our survey and compared these with objective life expectancies based on estimates of a mortality risk model that controls for a rich set of individual characteristics. We measured the respondents' knowledge on population life expectancy using two survey questions that provide information on whether the respondent has any idea what age people of his or her age and gender reach on average and, if so, to what extent his or her believes about remaining life expectancy are in line with age and gender specific population life expectancies from period life tables. This latter information on population life expectancy is related to the difference between respondent's predicted objective and subjective life expectancy.

The paper is structured as follows: Section 2 describes the data. Section 3 outlines the subjective and objective mortality risk models. Section 4 presents the estimation results and Section 5 concludes.

2. Data

We have used survey data from the Netherlands that have been supplemented with individual-level administrative data on dates of death. The subjective survival probabilities (SSPs) have been taken from the 1995 and 1996 waves of the DNB Household Survey (DHS) and we refer to Alessie et al. (2002) for a more general and detailed description of the DHS. These two waves of the DHS contain relatively many high-income households. Hence, while this does not invalidate our empirical findings it does suggest caution when extending our conclusions to the Dutch population. The data on the actual mortality of survey respondents are obtained from the Dutch causes of death registry (DO, DoodsOorzaken) 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 data contains a personal identifier which allowed us to determine 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.

We measure subjective life expectancy with individuals' answers to survival probability questions (Manski 2004). SSPs are elicited in the DHS 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. Respondents aged 25 through 65 report their probability of survival to age 75 and age 80; those aged 65 through 70, their survival expectations to age 80 and age 85; and respondents aged 70–75, 75–80, and 80–85, their expected survival probabilities to 85 and 90, 90 and 95, 95 and 100, respectively. 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, individuals' responses can be interpreted as survival probabilities conditional on them having reached their current age. For computational reasons when modeling mortality risk in section 3, we follow Perozek (2008) and have replaced reported probabilities of 0 and 1 by 0.01 and 0.99, respectively. [2]

[2] Our main findings remained when we either replace the probabilities 0 and 1 with 0.05 and 0.95, respectively, or by excluding 0 and 1 answers from the analysis.

The DHS survey has two questions about average life expectancy in the Netherlands. The first question is:

For people of your age and sex there is an average life expectancy. Do you have any idea what age people of your age and sex reach on average?

The answer to this question can be either yes or no. To respondents who answered yes, the following question is asked to elicit their beliefs about population life expectancy:

What age do you think people of your age and sex reach on average?

The reported ages in this question are compared to the corresponding population life expectancy from the age and gender specific 1995 life table available in the Human Mortality Database (HMD). We create two variables based on the responses given to these two questions. The variable 'Does not know population life expectancy' takes one if someone said no in the first question about average life expectancy. The variable 'Reported minus life table life expectancy' is defined as the difference between someone's beliefs about population life expectancy in the second question and his/her actuarial life expectancy. The latter is missing for the individuals who said no in the first question and, therefore, we assign zero to this difference if individuals have no idea about the average life expectancy in the Netherlands and control for whether or not they know population life expectancy in our analysis.

We also measure respondents' knowledge of the laws of probability using the following questions about income expectations:

What do you think is the probability that the total net income of your household will be less than €[LOWEST+(HIGHEST-LOWEST){0.2/0.4/0.6/0.8}] in the next 12 months?,* where LOWEST and HIGHEST are the minimum and maximum values of total net expected household income reported by the respondents, respectively. Respondents receive four questions consecutively where they are asked to report the probability that their total net annual household income will be less than 20/40/60/80 percent above their lowest expected yearly income. If respondents' answers satisfy the laws of probability we expect to find that these probabilities are increasing with the threshold level, that is, if respondents' income is less than 20 percent above their lowest expected income it is also less than 40 percent above their lowest expected income. We create a dummy variable 'Not able to answer probabilistic question on future income' which takes one if the respondent's answer does not satisfy this property. The definitions of all other variables used in our analysis are listed in the Appendix Table A1.

2.1. Sample Selection

The questions about population life expectancy were only asked to individuals who were either heads of household, spouses or cohabiting partners. We exclude individuals under 25 as many are still enrolled in education. If respondents were observed in both the 1995 and 1996 waves, we have used only the earlier response to avoid the potential influence of repeated interviewing on respondent behavior (Lazarsfeld 1940; Sturgis *et al.* 2009). After removing observations with missing information on the covariates for our analysis yielded a final sample of 2,245 individuals.

We have tested if individuals in our final sample have the same mortality risk as those excluded from the estimations. The *p-value* corresponding to the hypothesis test of equal mortality risks for the two groups was equal to 0.128 which suggests that our sample selection is not endogenous with respect to mortality.

2.2. Descriptive Statistics

For the analysis of subjective mortality risk we make use of two SSPs for each individual. Not all respondents provide two SSPs that can be used in the analysis. For instance, if a respondent answers that his survival probability to age 75 is less than or equal to his survival probability to age 80, this violates, what we refer to in this paper as, the strict monotonicity assumption (i.e. the survival up to age 75 should be larger than the survival up to age 80). According to Table 1, 601 respondents in our sample provided answers that violate the strict monotonicity assumption. Among these respondents 584 of them provided equal survival probabilities for two target ages (not reported in Table 1).

'Insert Table 1 about here'

Table 1 shows respondent characteristics for two samples: a sample with 1,644 respondents who provide strictly monotonic SSPs and a sample with 601 respondents who provided non-strictly monotonic SSPs. The table shows that compared to people who do not provide strictly monotonic SSPs, those who provide strictly monotonic SSPs more frequently report to have knowledge about population life expectancy, are more frequently able to answer probabilistic questions on future income, are more likely to be high educated and have high income, and are in relatively better health. The difference between beliefs about population life expectancy and actuarial life expectancy is, on average, -2.56 years for the individuals who provide strictly monotonic SSPs whereas it is -2.41 years for the individuals who provide non-strictly monotonic SSPs (bottom rows of the table). We have calculated the difference between beliefs about population life expectancy and actuarial life expectancy using period life tables. It is known that period life tables tend to underestimate life expectancy compared to cohort life tables when mortality rates decline over time (Hurd and McGarry 1995, 2002). As a result, we would expect the difference to be more negative if we used cohort life tables rather than period life tables. In other words, on average, the underestimation of population life expectancy seems to have little to do with the difference between period and cohort life tables.

Finally, Table 1 (bottom) shows the on average underestimation of life expectancy by respondents as also reported in previous studies: the average computed objective median life duration is around 85 years whereas the average subjective median life duration based on subjective survival probabilities is four year less and about 81 years.

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). Using a parametric model makes it possible to compare predicted life durations based on the subjective and objective mortality models for groups of individuals who differ with respect to observed characteristics such as their knowledge on population life expectancy.

3.1. Objective mortality model

T denotes a random variable representing the respondent's age at death. Respondent i with characteristics $\mathbf{x_i}$ is t_{0i} years old at the time of the survey (current age) and the probability of the respondents' age at death being greater than t is given by:

$$S(t|t_{0i},\mathbf{x_i};\gamma_O,\beta_O)=\Pr\left(T>t|t_{0i},\mathbf{x_i};\gamma_O,\beta_O\right)=\exp\left\{-\int_{t_{0i}}^{t}\theta(s|\mathbf{x_i};\gamma_O,\beta_O)ds\right\}, \qquad (1)$$

where $\theta(t|\mathbf{x_i};\gamma_O,\boldsymbol{\beta_O})$ is a mortality risk function and the assumption that T follows a Gompertz distribution together with a proportional hazard rate specification for the inclusion of individual characteristics yields to following functional form:

$$\theta\left(t|\mathbf{x_i};\gamma_O,\beta_O\right)=\exp\left(\gamma_O t+\mathbf{x_i}\boldsymbol{\beta_O}\right), \qquad (2)$$

where the parameter γ_O determined the age gradient in mortality and the parameter vector $\boldsymbol{\beta_O}$ determines how the various characteristics are related to mortality risk. The density function of dying at age t_i is given by $S\left(t|t_{0i},\mathbf{x_i};\gamma_O,\beta_O\right)\theta\left(t_i|\mathbf{x_i};\gamma_O,\beta_O\right)$ and the probability of a respondent still being alive at the end of observation period (December 2010) is given by the survivor function $S\left(t|t_{0i},\mathbf{x_i};\gamma_O,\beta_O\right)$. Based on these ingredients we obtained the estimates $\boldsymbol{\beta_O}$ and γ_O using Maximum Likelihood (Lancaster, 1990).

3.2. Subjective mortality model

As discussed, about 33 percent of respondents do not provide strictly monotonic SSPs and such answers cannot be used in the analysis. While one can argue that equal SSPs are related to rounding, this would require making additional assumptions to incorporate such observations in an empirical analysis (see, e.g., Perozek 2008 and Kutlu-Koc and Kalwij 2017). An alternative is to drop such observations and estimate a subjective mortality risk model using only individuals who provided answers that do not violate strict monotonicity. This latter approach is only valid when this would not yield a selection bias. A selection bias might, however, occur if those who were unable to provide strictly monotonic SSPs have different subjective mortality risks. For instance, Van Santen et al. (2012) who analyze responses to subjective retirement income replacement rate expectations show that using probabilistic survey questions yields an endogenous sample selection in the sense that removing responses that do not satisfy monotonicity yields a sample of people with more pessimistic expectations about the replacement rate. We, therefore, choose not to drop respondents who provide answers that violated strict monotonicity and employ an endogenous selection model in the spirit of Van Santen et al. (2012) and Heckman (1979).

The first part of the model is the selection equation:

$$C_i = \begin{cases} 0 & if \quad u_i \leq -\mathbf{w_i}\boldsymbol{\alpha} \\ 1 & if \quad u_i > -\mathbf{w_i}\boldsymbol{\alpha} \end{cases}, \tag{3}$$

where C_i is equal to one if the respondent provided (two) strictly monotonic SSPs, and zero otherwise. A vector of individual characteristic is denoted by $\mathbf{w_i}$. We assume u_i is normally distributed. For those who provide (two) strictly monotonic SSPs, we estimate the parameters of subjective mortality model using nonlinear least squares (NLLS) as follows:

$$SSP_{ji} = S_{ji}\left(t_{0i}, t_{ji}; \gamma_S, \lambda_S\right) + \rho IMR_i\left(\mathbf{w_i}\hat{\boldsymbol{\alpha}}\right) + \varepsilon_{ji} , \qquad j \in \{1,2\}, \quad i = 1,..., N , \tag{4}$$

with $S_{ji}\left(t_{0i}, t_{ji}; \gamma_S, \lambda_S\right) = \exp\left\{\frac{\lambda_S}{\gamma_S}\left(\exp\{\gamma_S t_{0i}\} - \exp\{\gamma_S t_{ji}\}\right)\right\}$ and $\lambda_S = \exp\{\mathbf{x_i}\boldsymbol{\beta}_S\}$, where SSP_{ji} are subjective survival probabilities and t_{ji} are the corresponding target ages, t_{0i} is the age at the time of the survey and $\varepsilon_{j,i}$ is an error term that is identically (also across *j*) and normally distributed. The above parametric function is equivalent with assuming a Gompertz distribution for life duration in which characteristics proportionally affect mortality risk (Kutlu-Koc and Kalwij 2017). Endogenous sample selection is controlled for in equation (4) by the inclusion of an inverse Mills ratio (IMR_i). Equations (3) and (4) are estimated in two steps. In the first step the estimates of $\boldsymbol{\alpha}$ are obtained from a Probit estimation and used to calculate the inverse Mills ratio that has been included in equation (4). In the second step, equation (4) has been estimated using the Non-Linear Least Squares.

3.3 Empirical specifications and model identification

In the models outlined above, the row vectors $\mathbf{x_i}$ and $\mathbf{w_i}$ contain the variables related to knowledge on population life expectancy, gender, educational attainment, household income, marital status, self-assessed health status and whether or not the individual has a chronic illness. We control as well for mood-effect when answering SSPs using reported happiness at the time of the survey (variable "happy"). In addition, we control for behavioral health risk using variables related to smoking, alcohol consumption and body mass index (BMI). The definitions of all variables are in Appendix Table A1.

The variable 'not able to answer probabilistic questions on future income' (see section 2) is used to identify our endogenous selection model of section 3.2. This variable can be regarded a proxy for knowledge on the laws of probability and is included in $\mathbf{w_i}$ but excluded from $\mathbf{x_i}$. With this so-called exclusion restriction we assume that the ability to answer probabilistic questions on future income does not affect subjective mortality risk (conditional on having controlled for $\mathbf{x_i}$).

4. Empirical Results

Table 2 show the estimation results of Equation (3) that relates a binary indicator of being able to provide strictly monotonic SSPs to variables measuring respondents' knowledge on population life expectancy, being unable to answer probabilistic questions on future income, and the other individual characteristics.

'Insert Table 2 here'

According to this table, the coefficient on being unable to answer probabilistic questions on future income is negative and statistically significant, suggesting that those who cannot answer the probabilistic questions on income consistently are less likely to provide strictly monotonic SSPs than those who are able to answer the income questions consistently. This finding shows that our exclusion restriction needed for model identification is relevant. The ability to answer SSPs correctly is not significantly related to having any idea about population life expectancy. Similarly, the difference between respondents' beliefs about population life expectancy and the actuarial life expectancy is not significantly associated with the ability to answer SSPs correctly. Overall, the *p-value* of the Wald test for knowledge indicates as well that the coefficients of two variables measuring respondents' knowledge on population life expectancy are not jointly significant at a five percent level of significance.

This table also shows that the higher educated are more likely to provide strictly monotonic SSPs than medium educated. On average, a high educated respondent is about 8 percentage point more likely to provide strictly monotonic SSPs compared to a medium educated person. Compared to the average rate of providing strictly monotonic SSPs in the sample which is about 73 percent, the size of the marginal effect is not negligible. Moreover, the *p-value* of the Wald test at the bottom of the table shows that the coefficients of low and high education are jointly significant at a one percent level of significance. The results furthermore show that respondents who are in good health are significantly less likely to provide strictly monotonic SSPs compared to respondents who are not in good health. Most objective health indicators, however, are insignificantly associated with the probability to provide strictly monotonic SSPs.

Table 3 reports the estimation results for the subjective mortality model with and without controlling for endogenous sample selection, and for the objective mortality model. The coefficient on the estimated inverse Mills ratio in column (2) is only statistically significant at a 10 percent significance level which suggests that endogenous sample selection is not a major concern in our subjective mortality risk model. The estimation results under columns (2) and (3) show as well that the estimates of the subjective mortality model without controlling for sample selection are rather similar to the ones after controlling for sample selection. These findings suggest that the removal of individuals who do not provide strictly monotonic SSPs from the estimation does not lead to inconsistent coefficient estimates of the subjective mortality risk model.

‘Insert Table 3 here’

According to the results in column (1), knowledge on population life expectancy does not play a role in explaining objective (actual) mortality risk within the sample. Based on the subjective mortality model (column (2)), respondents who do not know population life expectancy have significantly higher subjective mortality risk compared to those who have an idea about population life expectancy. The coefficient on the difference between respondents’ beliefs about population life expectancy and actuarial life expectancy is negative and statistically significant at a 1 percent level of significance. The subjective mortality risk is lower when the reported population life expectancy is greater than actuarial life expectancy, so individuals who overestimate the population life expectancy also report higher probabilities of survival. Conversely, the subjective mortality risk is higher when the difference is negative suggesting that individuals who underestimate population life expectancy also give lower probabilities of survival. These findings indicate that having knowledge on population life expectancy plays a significant role in answering questions about survival probabilities.

The coefficient on being happy in the objective model is statistically insignificant whereas it is significant at one percent in the subjective model, suggesting that individuals' answers to SSPs may be influenced by their mood, that is, they may think that they will live longer when they feel happier.

Table 3 also shows that women have significantly lower objective and subjective mortality risks than men but this gender difference is much smaller in the subjective model. Less educated people have significantly lower mortality risks than medium educated people. On the other hand, the coefficient on low education in the objective mortality model suggests that low education live shorter than medium educated although this difference is not statistically significant. Finally, the coefficients on smoking and obesity in the subjective model are smaller than those in the objective model, indicating that individuals underestimate the risk from smoking and obesity, as also found in Kutlu-Koc and Kalwij (2017).

Table 4 shows the extent to which individuals' knowledge on the population life expectancy explains the difference between subjective and objective remaining life expectancy. For this purpose, we have used the coefficient estimates in columns (1) and (2) of Table 3 to estimate the predicted median remaining life duration implied by objective and subjective mortality models for groups of individuals with different characteristics. As a reference individual (or group) we take a 50-year old married man, born in 1945, living in a middle income household, who knows population life expectancy about right, is happy, in good health, a non-smoker and non-drinker, and has no chronic illnesses, and has normal weight and a medium level of education. According to the results in Table 4, for this reference 50-year old man the difference between subjective and objective predicted life durations is -1.10 years and not statistically significant. Next, we change one characteristic of this reference individual at a time and these results are reported in the remaining rows. For instance, this table shows a difference of -5.55 years for a 50-year old woman which implies that women, on average, underestimate their remaining life expectancy compared to what is implied by the objective mortality model.

'Insert Table 4 here'

To test whether the difference between subjective and objective predictions is the same across characteristics we compare differences for each characteristic with the difference for the reference group (first row) and these are reported in the fifth column. The difference between 50-year old women and men suggests that women, on average, significantly underestimate their remaining life expectancy with about 4.45 years more than men. In line with Kutlu-Koc and Kalwij (2017), the table shows that male smokers significantly overestimate their remaining life expectancy with about 3.4 years compared to male non-smokers. Similarly, having chronic illnesses is associated with an overestimation of remaining life expectancy of three years. We also find that low educated overestimate and high educated underestimate their remaining life expectancy compared to medium educated, respectively, suggesting that individuals' level of education does not help them predict their remaining life expectancy more accurately. Predictions regarding marital status indicate that divorced men significantly overestimate their remaining life expectancy compared to married men.

The reference group in Table 4 are men who know population life expectancy about right in the sense that their beliefs about population life expectancy coincides with the life expectancy obtained from the life table. The difference between beliefs about population life expectancy and actuarial life expectancy is set equal to one for men who know population life expectancy but overestimates it with one year, and it is set equal to minus when they underestimate it with one year. According to the predictions in this table, men who do not know population life expectancy do not significantly underestimate their remaining life expectancy (third row, third column). On the other hand, these men underestimate their remaining life expectancy more than men who know population life expectancy about right (the reference) with a statistically significant difference of 2.35 years (third row, fifth column).

Table 4 also shows that men who know population life expectancy but underestimate it compared to actuarial life expectancy also underestimate their remaining life expectancy by 1.61 years, although this difference is not statistically significant. This finding is in line with Steffen (2009) who found that German men and women who underestimate population life expectancy also, on average, underestimate their own life expectancy, when compared to actuarial life expectancy. Moreover, we also find that the level of underestimation by men who underestimate population life expectancy is 0.51 years higher than the level of underestimation by men who know population life expectancy about right and this difference is statistically significant at one percent.

5. Conclusion

Our empirical analysis has demonstrated a significant role of having knowledge on population life expectancy for being able to answer probabilistic survival questions accurately: a one-year more accurate knowledge of population life expectancy is associated with an increased accuracy of about half a year in predicting one's own remaining lifetime. This suggests that a policy of informing individuals about population remaining life expectancies by age and gender may yield more accurate survival beliefs and, consequently, could help them making better decisions that require knowledge about their remaining life expectancy.

Acknowledgments

We wish to thank CentERdata for making the data available to us, Rob Alessie, Clemens Kool, Bertrand Melenberg, and seminar participants at the Netspar International Pension workshop (Amsterdam, January 2014), MEA workshop (Munich, June 2014) and European Society of Population Economics (Basel 2015) for valuable comments and discussions. Financial support has been provided by the Network for Studies on Pensions, Aging and Retirement (Netspar).

References

Alessie, R., Hochguertel, S. & van Soest, A. (2002). Household portfolios in the Netherlands. In L. Guiso, M. Haliassos, & T. Jappelli (Ed.), *Household Portfolios*, pp. 341–388. Cambridge, MA: MIT Press.

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.

Elder, T. (2013). The predictive validity of subjective mortality expectations: evidence from the health and retirement study. *Demography*, 50(2), 569-589.

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.

Groneck, M., Ludwig, A. and Zimper, A. 2013. "A life-cycle model with ambiguous survival beliefs," *Netspar Discussion Papers, DP10/2011-044*.

Heckman, J.J. 1979. "Sample Selection bias as a specification error," *Econometrica*, 47: 153–161.

Hurd, Michael D. 1989. Mortality Risk and Bequests. Econometrica 57 (4):779-813.

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.

Kutlu-Koc, V. , Alessie, R, and Kalwij, A., 2014. "Consumption behavior, annuity income and

mortality risk of the elderly", *Discussion Paper Series*, No. 12/2014-062, Network for Studies on Pensions, Aging and Retirement (Netspar).

Kutlu-Koc, V. and Kalwij, A., 2017. "Individuals' Survival Expectations and Actual Mortality," *European of Population*, forthcoming.

Lancaster, T. 1990. *The econometric analysis of transition data*. Econometric Society Monographs, Cambridge University Press.

Law, A., Kelton, W. 1982. "Simulation Modelling and Analysis," New York: McGraw-Hill.

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

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

McFadden, D., N. Schwarz, and J.Winter. 2004. “Measuring Perceptions and Behavior in Household Surveys,” paper presented at the Annual Meeting of the American Economic Association, San Diego, CA.

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

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

Van Santen, P., Alessie, R., and A. Kalwij, 2012. “Probabilistic Survey Questions and Incorrect Answers: Retirement Income Replacement Rates”, *Journal of Economic Behavior and Organization*, Volume 82, Issue 1, pp. 267-280

Salm, M. (2010). ‘Subjective mortality expectations and consumption and savings behaviours among the elderly,’ *Canadian Journal of Economics*, Vol. 43(3), pp. 1040–1057.

Steffen, B. 2009. “Formation and updating of subjective life expectancy: evidence from Germany,” *MEA Studies* 08.

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.

Appendix

Table A1: Variable definitions

Variable	Description	Variable
Happy	Based on the question *'All in all, to what extent do you consider yourself a happy person?'*	Dummy variable equal to 1 if very happy or happy, and 0 otherwise.
Woman	Respondent is a woman	Dummy variable equal to 1 if woman, 0 if man
Low education	Primary (or lower) education or vocational training through an apprentice system	Dummy variable equal to 1 if lower educated, 0 otherwise
High education	Vocational college or university degree	Dummy variable equal to 1 if higher educated, 0 otherwise
Income	The sum of the net annual incomes of all household members after deduction of taxes but before making payments such as rent, mortgages, etc.	A continuous variable in Dutch guilders which is standardized using the equivalence scale provided by Statistics Netherlands (Siermann et al., 2004). Used to construct the variables "Low income" and " High income"
Low income	Defined as income in the first tertile of the distribution	Dummy variable equal to 1 if the household has a low income, 0 otherwise
High income	Defined as income in the third tertile of the distribution	Dummy variable equal to 1 if the household has a high income, 0 otherwise
Good health	Self-reported health (excellent, good, fair, not so good, poor)	Dummy variable equal to 1 if the person in excellent or good health, 0 otherwise
No Smoking	Never smoked, currently smoking or stopped smoking	Dummy variable equal to 1 if the person never smoked, 0 otherwise
Alcohol	Alcohol consumption measured in number of drinks a day	Dummy variable equal to 1 if consumes more than 4 alcoholic drinks a day, 0 otherwise
Chronic illness	Long- term illness, disorder, disability, consequences of accident	Dummy variable equal to 1 if has a chronic illness, 0 otherwise
Overweight	$25 \leq$ Respondent's body mass index (BMI) < 30	Dummy variable equal to 1 if ($25 \leq$ BMI < 30), 0 otherwise
Obese	Respondent's BMI ≥ 30	Dummy variable equal to 1 if BMI ≥ 30, 0 otherwise
Birth year	Respondent's year of birth	A continuous variable
Single	Respondent is single	Dummy variable equal to 1 if single, 0 otherwise
Widowed	Respondent is widowed	Dummy variable equal to 1 if married, 0 otherwise
Age	Respondent's age at the time of interview	A continuous variable with an accuracy in months

Table 1: Summary statistics for those who provided SSPs that satisfied strict monotonicity and those who did not. SSP=Subjective Survival Probability.

	SSPs that satisfied strict monotonicity	SSPs that did not satisfy strict monotonicity
Number of individuals	1,644	601
Variables	Proportion in %	Proportion in %
Does not know population life expectancy	20.19	26.29
Not able to answer probabilistic questions on future income	58.64	67.72
Happy	86.5	87.85
Female	36.25	41.76
Low education	18.67	23.46
Medium education	32.24	39.6
High education	49.09	36.94
Low income	33.03	39.6
Medium income	27.8	26.46
High income	39.17	33.94
Married	87.77	88.35
Divorced	3.71	3.00
Widowed	2.01	2.33
Single	6.51	6.32
Chronic illness	23.24	24.29
No smoking	69.28	70.72
Alcohol	8.88	7.65
Good health	82.97	86.52
Overweight	34.49	36.44
Obese	5.17	4.83
Died before December 2010	11.98	8.82
	Average in years	Average in years
Age	47	47
Median life duration conditional on having died before December 2010 (250 individuals)	75.5	74.92
Average objective life expectancy, conditional on alive at current age, computed [a]	84.97	85.23
Average subjective life expectancy, conditional on alive at current age, computed [b]	80.69	-
Reported life table population minus life expectancy (conditional on knowing population life expectancy) [c]	-2.56	-2.08

Notes: [a] Based on the estimation results of a Gompertz mortality model with only gender, age, and the year of birth as covariates. [b] Based on SSPs and assuming a Gompertz mortality model. [c] Population life expectancy is based on period life tables.

Table 2: Estimation results of the selection equation (3). SSP =Subjective Survival Probability.

Dependent variable: Being able to provide SSPs that satistify strict monotonicity			
Covariates	Parameter estimate	Standard error	Marginal effect [a]
Does not know population life expectancy	-0.085	0.074	-0.028
Reported minus life table population life expectancy	-0.013	0.009	-0.004
Not able to answer probabilistic questions on future income	-0.208***	0.061	-
Happy	-0.034	0.089	-0.011
Female	-0.104	0.064	-0.034
Birth year	-0.014	0.051	-0.004
Age (in months and years)	-0.013	0.051	-0.004
Low education	-0.023	0.079	-0.007
High education	0.242***	0.068	0.078***
Low income	-0.053	0.075	-0.017
High income	0.031	0.074	0.01
Divorced	0.129	0.162	0.041
Widowed	-0.061	0.205	-0.02
Single	0.047	0.121	0.015
Chronic illness	-0.122	0.079	-0.041
No smoking	-0.059	0.064	-0.019
Alcohol	0.042	0.107	0.013
Good health	-0.271***	0.095	-0.083***
Overweight	-0.05	0.062	-0.016
Obese	0.054	0.135	0.017
Constant	30.372	101.406	
Number of observations	2,245		
p-value Wald test: knowledge	0.073		
p-value Wald test: education	0.000		
p-value Wald test: income	0.500		
p-value Wald test: marital status	0.831		
p-value Wald test: BMI	0.616		
Notes: * p < 0.10, ** p < 0.05, *** p < 0.01.			

[a] The marginal effect on the probability of being able to provide SSPs that satistify strict monotonicity.

Table 3: Estimation results for the objective and subjective mortality risk models

	Objective Mortality (1)	Subjective Mortality (Endogenous sample selection) (2)	Subjective Mortality (Exogenous sample selection) (3)
Does not know population life expectancy	-0.132	0.142***	0.123***
	(0.177)	(0.031)	(0.028)
Reported minus life table life expectancy	0.026	-0.026***	-0.028***
	(0.021)	(0.004)	(0.004)
Happy	-0.215	-0.131***	-0.137***
	(0.173)	(0.032)	(0.031)
Female	-0.737***	-0.110***	-0.133***
	(0.168)	(0.028)	(0.024)
Birth year	0.010	0.002**	0.003*
	(0.015)	(0.001)	(0.001)
Low education	0.131	-0.106***	-0.110***
	(0.165)	(0.032)	(0.032)
High education	-0.232	-0.018	0.028
	(0.158)	(0.033)	(0.025)
Low income	-0.046	0.005	-0.007
	(0.180)	(0.028)	(0.027)
High income	-0.084	0.002	0.009
	(0.170)	(0.028)	(0.026)
Divorced	0.415	-0.026	-0.002
	(0.277)	(0.051)	(0.056)
Widowed	0.244	0.076	0.063
	(0.268)	(0.105)	(0.077)
Single	-0.174	-0.075*	-0.068
	(0.306)	(0.043)	(0.044)
Chronic illness	0.476***	0.093***	0.071**
	(0.156)	(0.034)	(0.028)
No smoking	-0.632***	-0.141***	-0.151***
	(0.143)	(0.024)	(0.023)
Alcohol	0.345*	0.053	0.058
	(0.209)	(0.039)	(0.036)
Good health	-0.286*	-0.222***	-0.268***
	(0.167)	(0.042)	(0.031)
Overweight	-0.137	0.044*	0.036
	(0.139)	(0.024)	(0.023)
Obese	0.304	0.092*	0.104**
	(0.235)	(0.052)	(0.047)
Constant	-31.236	-15.137***	-14.826***
	(31.064)	(2.767)	(2.838)
Age	0.010***	0.008***	0.008***
	(0.001)	(0.000)	(0.000)
Inverse Mills Ratio	-	0.122*	-
		(0.066)	
Number of observations	2,245	2,245	1,644
p-value Wald test: knowledge	0.419	0.000	0.000
p-value Wald test: education	0.104	0.004	0.000
p-value Wald test: income	0.883	0.979	0.821
p-value Wald test: marital status	0.328	0.305	0.367
p-value Wald test: BMI	0.182	0.057	0.044

Notes: * p < 0.10, ** p < 0.05, *** p < 0.01. Column (2): bootstrapped standard errors are based on 1000 replications.

Table 4: Comparison of objective and subjective predicted life expectancy in years for various groups of people.

Remaining individual life expectancy	objective	subjective	Difference	std. error	Difference of differences	std. error
	(1)	(2)	(2)-(1)			
Man (reference) [a]	33.80	32.69	-1.10	2.19	0.00	-
Woman	39.63	34.09	-5.55*	2.90	-4.45***	1.28
Man does not know PLE	34.87	31.42	-3.45	2.43	-2.35**	1.14
Man knows PLE but underestimates with one year	34.01	32.40	-1.61	2.20	-0.51***	0.13
Man knows PLE but overestimates with one year	33.59	32.99	-0.60	2.18	0.50***	0.13
Unhappy man	32.13	31.27	-0.86	2.29	0.24	1.06
Smoking man	28.79	31.11	2.33	1.89	3.43***	1.02
Obese man	31.38	31.61	0.22	2.36	1.32	1.37
Man in bad health	31.49	29.89	-1.60	2.26	-0.50	1.04
Man drinking alcohol	31.09	32.09	1.00	2.24	2.10	1.33
Man with chronic illnesses	29.99	31.93	1.95	1.97	3.05***	0.99
Low educated man	32.71	33.84	1.13	2.19	2.23**	0.99
High educated man	35.61	32.38	-3.23	2.43	-2.13**	1.04
Man living in a low income household	34.10	32.76	-1.35	2.32	-0.25	1.16
Man living in a high income household	34.49	32.60	-1.89	2.26	-0.79	1.06
Divorced man	30.45	32.70	2.25	2.47	3.35**	1.68
Widowed man	31.86	32.02	0.17	2.48	1.27	1.37
Single man	35.25	33.43	-1.83	2.93	-0.73	1.96

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. [a] The reference is a 50-year old married man, born in 1945, living in a middle income household who knows
Objective and subjective life expectancy, and the differences, are obtained by Monte Carlo simulations using 1000 replications (Law and Kelton 1982).