Utrecht University

School of Economics

Is the Accuracy of Individuals' Survival Beliefs Associated with Their Knowledge of Population Life Expectancy?

Vesile Kutlu Koc, Adriaan Kalwij

Utrecht University

Is the Accuracy of Individuals' Survival Beliefs Associated with Their Knowledge of Population Life Expectancy?

Vesile Kutlu Koc
Munich Center for the Economics of Aging
Adriaan Kalwij
Utrecht School of Economics
Utrecht University

November 2020

Abstract
Previous studies have shown that, on average, individuals are pessimistic about their remaining lifetime, which could yield suboptimal long-term decisions. Using Dutch household survey data supplemented with death registry data, we found that individuals with a one-year better knowledge of population life expectancy had a significantly smaller difference of about 0.3 years, on average, between their predicted subjective and objective remaining lifetime. This finding was robust to whether socioeconomic status and health-related covariates were controlled for. Our findings may suggest that informing individuals about population life expectancies for people of their age and gender, can help them to more accurately predict their remaining lifetime.

Keywords: mortality risk, subjective survival, population life expectancy, decision making

JEL classification: C41, D8, I1

BACKGROUND

On average, individuals underpredict their remaining lifetime, which could yield suboptimal long-term decisions.

OBJECTIVE

Is individuals' under- or overprediction of remaining lifetime associated with their knowledge of population life expectancy of people of their age and gender?

METHODS

Mortality risk models are estimated using the 1995 and 1996 waves of the Dutch DNB Household Survey with data on individuals' subjective survival beliefs and their knowledge of population life expectancy. These data are supplemented with administrative data on mortality from the causes of death registry for the years 1995 to 2018.

RESULTS

Individuals with a one-year better knowledge of population life expectancy had a significantly smaller difference of about 0.3 years, on average, between their predicted subjective and objective remaining lifetime. This finding was robust to whether socioeconomic status and health-related covariates were controlled for.

CONTRIBUTION

We provide empirical evidence in support of the hypothesis that individuals with a better knowledge of population life expectancy predict more accurately their remaining lifetime.

1. Introduction

Individuals' long-term decisions on matters such as human capital investments, savings, the purchase of a life insurance policy, retirement timing, or adopting a healthy lifestyle can be influenced by their beliefs on remaining lifetime (Bissonnette et al. 2017; O'Dea and Sturrock 2019; Oster et al. 2013; Vanajan et al. 2020). For instance, theoretical economic models of life cycle behavior, such as Hurd (1989), have shown the importance of individual variation in mortality risk for savings behavior of elderly singles, and empirical support for this model has been presented in Gan et al. (2004), Salm (2010), and Kutlu Koc et al. (2017). At the same time, while subjective survival probabilities elicited in household surveys (Manski 2004) have been shown to predict actual survival (Hurd and McGarry 2002), several studies have shown that individuals on average underpredict their remaining lifetime (Groneck et al. 2016; Kutlu Koc and Kalwij 2017; Teppa 2012). Such underprediction, or overprediction for that matter, of remaining lifetime can lead to individuals making suboptimal long-term decisions. For instance, Bissonnette et al. (2017) showed that misperceptions of mortality risk can lead to substantial welfare losses, and Heimer et al. (2019) showed that pessimistic survival beliefs at younger ages and optimistic beliefs at older ages explain, respectively, under-saving before retirement and a slower rate of dissaving after retirement.

An important policy-relevant issue we addressed in our study is whether individuals' under- or overprediction of remaining lifetime is associated with their knowledge of population life expectancy of people of their age and gender. If so, informing individuals on population life expectancy could help them to more accurately predict their remaining lifetime. Two previous studies suggest that respondents' knowledge of population life expectancy could be associated

with 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 on population survival rates. Steffen (2009) combined the elicited beliefs of German respondents on population life expectancy and on their own position relative to the broader expectancy. 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 their actuarial life expectancy. These findings of Steffen (2009) suggest that individuals' underprediction of their own remaining lifetime can be related to their underestimation of population life expectancy generally. Both studies, however, did not provide empirical evidence on the relationship between individuals' knowledge of population life expectancy and the difference between their subjective and their objective survival rates.

Our main contribution to the literature is that we show that individuals who had better knowledge of population life expectancy predicted their remaining lifetime significantly more accurately. This may suggest that informing individuals about population life expectancies for people of their age and gender, could help them to more accurately predict their remaining lifetime and, arguably, to make better long-term decisions.

The paper is structured as follows: First we discuss in Section 2 our data that was taken from the 1995 and 1996 waves of the Dutch DNB Household Survey, and we supplemented it with administrative data on mortality from the causes of death registry for the years 1995 to 2018. We measured respondents' knowledge of population life expectancy using two survey questions that elicited whether they knew what age people of their own age and gender reached on average, and if so, what this age was. The difference between a respondent's belief regarding population life

expectancy and their actuarial life expectancy was taken as a measure of respondent's knowledge on (remaining) population life expectancy. Next, we quantify in Section 3 how the knowledge people have of population life expectancy relates to the accuracy of subjective survival beliefs. This accuracy was measured by calculating the differences between predicted subjective and objective remaining lifetimes that were based on estimates of mortality risk models using data on observed mortality, or using subjective survival probabilities. These models controlled for a rich set of individual characteristics related to respondents' socioeconomic and health status to address possible endogeneity of knowledge on population life expectancy as a covariate in such models. Section 4 concludes and discusses the main findings and short comings of our study, and how future research can address these.

2. The Data and Descriptive Statistics

The raw data of the 1995 and 1996 waves of the DNB Household Survey (DHS) contains information on 9,415 individuals from 3,348 households. The DHS oversampled high-income households, and while this does not invalidate our empirical findings, it warrants caution when extending our conclusions to the general Dutch population. We refer to Alessie et al. (2002) for a detailed description of the DHS. We have supplemented our survey data with administrative microdata from the causes of death registry that contains the year of death of Dutch residents who died during the 1995–2018 period (CBS 2020). The largest reduction in sample size was due to the selection of about 20% of respondents who were aged 52-84 in 1995 or aged 53-84 in 1996. These respondents were the only ones for whom we could determine whether they died before the target age for which they provided subjective survival probabilities (see below). Item non-response

caused a further sample reduction of about 30%. Our final estimation sample contained information on 1,273 respondents. Appendix A describes the data cleaning process in detail, provides definitions of all variables used for our analysis, and the variables' sample means and standard deviations.

Objective survival until a certain age is based on mortality information from the death registry that covered the years 1995-2018. Respondents were followed from the year of interview (1995 or 1996) until the end of 2018 or until their death (whichever came first). During the period 1995-2018, 629 respondents died and their year of death was observed. For our analysis we compared this objective information on individuals' survival with their beliefs on their survival probabilities to certain ages at the time they were surveyed (Manski 2004). These latter subjective survival probabilities (SSPs) were elicited in the DHS using the survey question

What do you think the chances are that you will live to be T years of age or more?

Here $T \in \{75, 80, 85, 90, 95, 100\}$ is a target age that depends on the respondent's age at the time of the survey. Respondents aged 52-64 reported their SSP to age 75 and age 80; those aged 65-69, their SSPs to age 80 and age 85; and those aged 70–74, 75–79, and 80–84, reported their SSPs to ages 85 and 90, 90 and 95, or 95 and 100, respectively. These responses were measured on a 10-point scale, from 0, "no chance at all," to 10, "absolutely certain." Following Hurd and McGarry (1995), we assumed that after having divided these responses by 10, they could be interpreted as survival probabilities conditional on individuals having reached their current age. Further, following Perozek (2008), we replaced reported probabilities of 0 and 1 by 0.01 and 0.99,

respectively, and when equal SSPs were reported, we added 0.05 to the SSP for the lowest target age and subtracted 0.05 from the SSP for the highest target age (about 30% of the cases).[1]

In line with the existing literature, Fig. 1 shows that respondents, on average, underpredicted their survival probabilities. For all target ages 75-100 together, respondents' beliefs on survival probabilities were, on average, 16 pp (percentage points) lower than the observed survival rate. We observed underprediction of survival probabilities for all target ages except for the highest target ages of 95 and 100.

< Fig.1 about here >

The DHS survey also asked two questions about population life expectancy in the Netherlands. The first question was:

> *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 variable DK-PLE takes the value one if the respondent answered no to this question, and zero otherwise. To respondents who answered they know population life expectancy (DK-PLE=0), the following question was asked to elicit their subjective knowledge of population life expectancy:

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

---

[1] Perozek (2008) performed sensitivity checks concerning these assumptions, Kleinjans and Van Soest (2014) provided evidence that taking into account rounding and 50% focal point, answers did not significantly affect the coefficient estimates of the determinants of SSPs, and the reliability of SSPs was confirmed by de Bresser (2019).

The reported ages in this second question were compared to the corresponding population life expectancy from age and gender specific actuarial cohort life tables (Royal Dutch Actuarial Association, 2019). These life tables contain actual mortality rates until 2012, and predicted mortality rates from 2012 until 2062. Based on these tables, all men in our sample were expected to live at least until the age of 81, while only 6% of the men in our sample, in answering the second question, gave 81 years or more as the population life expectancy (Fig. 2). This latter percentage equaled 8% when the DKs were excluded (when DK-PLE=1). Hence, and in line with the findings of Steffen (2009) for Germany, most men believed the population life expectancy for men of their age to be lower than the actuarial one. A similar pattern emerged for women: based on life tables, all women in our sample were expected to live until the ages 85-90, while only 6% believed this to be the case (8% when excluding the DKs). The aggregation over ages in Fig. 2 conforms to the publication requirements of Statistics Netherlands, and the disaggregated numbers were used for the further figures and the empirical analysis.

< Fig.2 about here >

For the empirical analysis we used the difference between the reported and actuarial (remaining) population life expectancy as a measure of respondents' knowledge of population life expectancy for people of their age and gender. This variable is referred to as PLE-knowledge (Fig. 3). The underestimation of population life expectancy was on average about six years. Not reported in this figure, is that less than 1% of the respondents had this age exactly right (PLE-knowledge=0).

For Fig. 4, Figs. 1 and 3 were combined to present *prima facie* evidence on the relationship between PLE-knowledge and individuals' accuracy regarding their SSPs. Only, for this figure we grouped PLE-knowledge into three categories (defined in the figure's footnote). The main insight

this figure represents is that respondents who had better knowledge of population life expectancy, on average provided more accurate SSPs. Respondents who answered they have no knowledge of population life expectancy (DK-PLE=1), on average underestimated their own survival probabilities with 16 pp; an underestimation which is about the same as for respondents who indicated they had knowledge of population life expectancy and underestimated it with six to eight years.

< Fig.3 about here >

< Fig.4 about here >

3. Empirical Analysis

Following previous empirical studies on individual mortality, we modelled remaining lifetimes with a Gompertz distribution (Gompertz 1825; Olshansky and Carnes 1997; Perozek 2008). Mortality risk models were estimated using objective information on mortality as observed over the period 1995-2018, and using SSPs. Appendix B provides modeling and estimation details. The objective and subjective mortality risk models controlled for the same individual characteristics such as gender, educational certification attained, and health behavior, all of which have been shown to be related to objective or subjective mortality (e.g., Cutler et al. 2006, 2011; Delavande and Rohwedder 2011; Hurd and McGarry 2002; Kalwij et al. 2013; Kutlu Koc and Kalwij 2017).

3.1 Main empirical findings

In line with previous findings, the predicted mortality rates by age based on the estimation results of the subjective mortality risk model are higher than those based on the results of the objective model, except for the highest age (Table 1, Panel A). In line with Kutlu Koc and Kalwij (2017) for the Netherlands, the estimated age gradient for the subjective relative to the objective model was steeper, and lower mortality among women than among men for the objective model was not found for the subjective model (Panel B). The results for all covariates included in our models are reported in Table C1 and are summarized in Panel C of Table 1. Not reported here, is that the relationships between the covariates for socioeconomic status and mortality risk were jointly significant for the objective model when health-related covariates were omitted.[2]

Knowledge of population life expectancy is (jointly) insignificantly related to mortality risk in the objective model (left columns of Panel B, Table 1), and significantly related to mortality risk in the subjective model (right columns). Individuals who did not know population life expectancy had, on average, about a 21% higher subjective mortality rate than individuals who knew their population life expectancy exactly. Compared to this latter group, individuals who reported a population life expectancy of one year lower than the correct one, had a 3% higher subjective mortality rate. In other words, the better individuals' knowledge on population life expectancy, the lower, on average, was their subjective mortality rate.

---

[2] This relates to using a relatively small sample. See Kutlu Koc and Kalwij (2017) for more details on the socioeconomic status and health associations with subjective or objective mortality when using a larger sample drawn from the same surveys that we used.

< Table 1 about here >

The predicted objective lifetime for a male reference individual is about 88 years, and his predicted subjective lifetime is almost seven years lower, at about 81 years (Table 2, Panel A). These predictions are conditional on having survived until the age of 55 and, given our model specification (Appendix B), the reported marginal changes in Panel B on the difference between the predicted subjective and objective remaining lifetimes are insensitive to the characterization of the reference individual. The degree of underpredicting remaining lifetime did not vary significantly with most individual characteristics, with only a few exceptions (Panel B). On average, women underpredicted their remaining lifetime with about five years more than men (the reference individual is male) and, on average, smokers and obese individuals underpredicted their remaining lifetime less than, respectively, non-smokers and individuals with normal weight (BMI<25). Individuals who reported to be in bad health underpredicted their remaining lifetime significantly more.

Individuals with better knowledge of population life expectancy had, on average, relatively more accurate beliefs regarding their remaining lifetimes (Table 3, top: main specification I). The reference group in Table 3 are individuals who answered they know population life expectancy, but had underestimated it with six years (sample average); individuals who answered they do not know population life expectancy, underpredicted remaining lifetime, on average, about the same as the reference group (top row). Individuals who knew population life expectancy exactly, underpredicted their remaining lifetime less than the reference group by, on average, predicting an almost two-years more accurate belief of their remaining lifetime. In terms of marginal changes, the last row for specification I shows that individuals who underestimated population life

expectancy with one year less, underpredicted their remaining lifetime with, on average, about 0.3 years less.

< Table 2 about here >

3.2 Sensitivity Analyses

Knowledge of population life expectancy is, arguably, related to characteristics that determine objective and subjective survival rates. Figure 4 also hinted at this, as the observed survival probability was, on average, relatively lower for those who answered they do not know population life expectancy (DK-PLE=1). To address this issue of possible endogeneity of DK-PLE and PLE-knowledge in our mortality risk models, we controlled for a rich set of socioeconomic and health characteristics, to mitigate possible endogeneity biases. For instance, one can argue that the estimated change in the accuracy of subjective remaining lifetime due to one-year change PLE-knowledge is not the marginal effect of knowledge of population life expectancy *per se,* but confounds the effect of unobserved factors on the accuracy. If so, and if these unobserved factors are correlated with our covariates, we would expect a bias in the marginal accuracy change in subjective remaining lifetime due to better knowledge of population life expectancy when not controlled for covariates. As shown in specification II of Table 3, excluding socioeconomic and health characteristics resulted in an effect of PLE-knowledge that is rather similar to the one of specification I, which is in support of exogeneity of PLE-knowledge. As always, one can still question the exogeneity of knowledge of population life expectancy in our mortality models, but the empirical evidence presented here provided no arguments for it.

We carried out robustness checks with respect to how PLE-knowledge entered the mortality models, and considering the possible influences of outliers. In specification I, the relatively marginal change in mortality risk due to a change in knowledge of population life expectancy is the same across the range of PLE-knowledge (Eq. (B1), Appendix B). Specification III, Table 3, used, instead, a spline function to allow this relative marginal change to vary with PLE-knowledge. For instance, the more population life expectancy was underestimated, the larger this change could be. As the results show, using a spline function did not affect the main findings of specification I. Not reported here, is that the estimated coefficient of a quadratic term, when using a quadratic instead of a linear function of PLE-knowledge, was not significant. Removing potential outliers[3] or removing individuals who did not know population life expectancy (DK-PLE=1) also did not change our main findings (specifications IV and V).

Finally, specifications VI and VII of Table 3 present the main findings of specification I by gender. For men, the main findings remained, but for women they turned out to be insignificant. Results not reported on in the table, revealed that the statistically insignificant findings for women are due to a large and insignificant coefficient of PLE-knowledge in the objective mortality model (a marginal change of -1.6% with a standard error of 2.7)[4], while it is strong and significant in the subjective mortality model: a one-year smaller underestimation of PLE reduced the subjective mortality rate with 2%. This latter percentage is somewhat smaller than for specification I (Table C1), but together with the insignificant relationship between PLE-knowledge and the objective

---

[3] This removal included the few individuals who overestimated population life expectancy.

[4] Arguably, sample size, or rather the number of deaths, mattered for this imprecision of the estimates of the objective mortality model (551 women in our sample, of whom 228 died before the end of 2018).

mortality rate, it is in favor of women with better knowledge of population life expectancy predicting their remaining lifetime more accurately.

< Table 3 about here >

4. Discussion

In line with the findings of previous studies, we found that individuals, on average, underpredicted their remaining lifetime, and that this underprediction varied with gender, age, and health-related characteristics. Another strand of literature has shown that individuals make long-term decisions regarding matters such as savings, purchasing a life insurance policy, or adopting a healthy lifestyle, based on beliefs about their remaining lifetime. Arguably, individuals could make better long-term decisions if they were to have more accurate beliefs about their remaining lifetime. This study adds to these strands of literature by showing that individuals with better knowledge of population life expectancy have more accurate beliefs about their remaining lifetime: respondents with a one-year better knowledge of population life expectancy, had, on average, a significantly smaller difference between predicted subjective and objective remaining lifetimes of about 0.3 years.

Our findings suggest, arguably, that informing individuals about population life expectancies, for people of similar age and gender, could lead to more accurate survival beliefs. Such information on population life expectancy can, for instance, be provided together with the information individuals often receive annually about their accrued public pension entitlements. Notably, and related to the latter suggestion, the increased population life expectancy over the last few decades

has been at the center of many recent public debates on pension reforms in many countries, including the Netherlands. The increased population life expectancy is often directly linked to governments' plans to increase the state pension age. It is, therefore, likely that individuals currently are better informed on remaining life expectancy than our respondents who were interviewed in the mid-1990s. Future research can provide insight into whether individuals nowadays have more accurate survival beliefs than they had in the mid-1990s, and if so, to what extent this can be attributed to better knowledge of population life expectancy.

Finally, knowledge of population life expectancy is arguably exogenous in our mortality risk models (see Section 3.2) but there are always arguments for this variable to be endogenous. For instance, unobserved health conditions may prompt individuals to inquire about population life expectancy, or (unobserved) cognition can be correlated with knowledge of population life expectancy and (subjective) mortality, after having controlled for a rich set of observed covariates. The need for more conclusive evidence on this issue of endogeneity is, therefore, apparent. For instance, future research can use an experimental setup, randomly providing some survey participants with information on population life expectancy and then eliciting, before and after providing the information, their knowledge of population life expectancy, as well as their subjective survival probabilities. In such a setup, subsequently investigating the accuracy of these survival beliefs is most likely not feasible in the short term, as it would require following respondents for many years. In the absence of such experimental evidence, to our knowledge, this study, to date, is the first to provide empirical evidence in support of the hypothesis that individuals with a better knowledge of population life expectancy predict more accurately their remaining lifetime.

Acknowledgments

References

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

Bissonnette, L., Hurd, M. D., and Michaud, P.-C. (2017). Individual survival curves comparing subjective and observed mortality risks. *Health economics*, 26(12):285-303.

de Bresser, J. (2019). Measuring subjective survival expectations – Do response scales matter? *Journal of Economic Behavior & Organization*, 165: 136-156.

CBS (2020). *Microdata Services.* www.cbs.nl/microdata

Cox, D.R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34 (2):187-220.

Cutler, D.M., Deaton, A.S. & Lleras-Muney, A. (2006). Determinants of Mortality. *Journal of Economic Perspectives*, 20(3): 97-120.

Cutler, D.M., Lange F., Meara E., Richards-Shubik S. & Ruhm C.J. (2011). Rising educational gradients in mortality: The role of behavioral risk factors. *Journal of Health Economics*, 30: 1174-1187.

Delavande, A. & Rohwedder, S. (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.

Froot, K. A. (1989). Consistent covariance matrix estimation with cross-sectional dependence and heteroskedasticity in financial data. *Journal of Financial and Quantitative Analysis*, 24: 333–355.

Gan, L., Gong, G., Hurd, M. & 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. & Zimper, A. (2016). A life-cycle model with ambiguous survival beliefs. *Journal of Economic Theory*, 162: 137-180.

Heimer, R.Z., Myrseth, K.O.R. & Schoenle, R.S. (2019). YOLO: Mortality Beliefs and Household Finance Puzzles. *Journal of Finance*, 74(6): 2957-2996

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

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

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

Kleinjans, C. and Van Soest, A. (2014). Non-response and focal point answers to subjective probability questions. *Journal of Applied Econometrics*, 29(4): 567-585.

Kalwij, A., Alessie, R. & Knoef, M. (2013). Individual Income and Remaining Life Expectancy at the Statutory Retirement Age of 65 in the Netherlands. *Demography*, 50(1): 181-206

Kutlu Koc, V., Alessie, R. & Kalwij, A. (2017). Consumption Behavior, Annuity Income and Mortality Risk of Retirees. *De Economist*, 165 (3): 349-380.

Kutlu Koc, V. & Kalwij, A. (2017). Individual Survival Expectations and Actual Mortality: Evidence from Dutch Survey and Administrative Data. *European Journal of Population*, 33 (4): 509-532.

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

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

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

O'Dea, C. & Sturrock, D. (2019). Survival pessimism and the demand for annuities. *IFS Working Paper*, W19/02, Institute for Fiscal Studies

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

Oster, E., Shoulson. I., & Dorsey, E.R. (2013). Limited Life Expectancy, Human Capital and Health Investments. *American Economic Review*, 103 (5): 1977-2002.

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

Royal Dutch Actuarial Association. 2019. Available from: https://www.ag-ai.nl/

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

Steffen, B. (2009). Formation and updating of subjective life expectancy: evidence from Germany. *MEA Studies* 08. he Munich Center for the Economics of Aging, Germany.

Sturgis, P., Allum, N. & Brunton-Smith, I. (2009). Attitudes over time: The psychology of panel conditioning. In *Methodology of Longitudinal Surveys*, edited by P. Lynn: 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, De Nederlandsche Bank, the Netherlands

Vanajan, A., Bültmann, U. & Henkens, K. (2020). Why do older workers with chronic health conditions prefer to retire early? *Age and Ageing*, doi: 10.1093/ageing/afz180

White, H. (1980). A Heteroscedasticity-Consistent Covariance Matrix and a Direct Test for Heteroscedasticity. *Econometrica,* 48: 817-838

Appendix A   Data Cleaning and Definitions of the Variables

The raw DHS data contains information on 7,505 individuals in 1995 and 6,900 individuals in 1996 (9,415 unique individuals from 3,348 households). For 5.6% of these individuals the personal identifier to link it with the death registry was missing, and 1.8% had emigrated after being interviewed. For these respondents we could not establish whether they had died during the 1995-2018 period, therefore they were removed from the sample. For respondents who were in both the 1995 and 1996 waves, we have used the 1995 responses to avoid a potential influence of repeated interviewing on response behavior (Lazarsfeld 1940; Sturgis et al. 2009). Our baseline sample, therefore, consisted of 6,906 individuals interviewed in 1995 and 1,804 individuals interviewed in 1996. We selected 1,755 respondents aged 52-84 in 1995 or 53-84 in 1996, as these were the only respondents for whom we could determine whether or not they died before the target age for which they provided subjective survival probabilities (see main text). We removed 139 respondents who did not answer the questions on population life expectancy, nine respondents who reported survival probabilities that were increasing with an increasing target age, and five respondents who did not answer the subjective survival probability questions. After removing a further 329 respondents due to missing values on the other variables selected for the analysis, our final sample consisted of 1,273 respondents from 889 households. This further cleaning after the age-selection was not significantly related to mortality risk (a p-value of 0.17), which suggests exogenous sample selection with respect to mortality risk. About half of the respondents (629) died before the end of 2018 and for these respondents the year of death was observed. The definitions of the variables for our analysis, and their sample averages and standard deviations, are listed in Table A1.

Table A1   Sample means, standard deviations, and definitions of all variables

Variable	N	Mean	Std.Dev.	Definition
Year 1996	1,273	0.09	0.29	Equal to one for 1996, 0 for 1995
Birthyear	1,273	1933	7.15	Year of birth
Age	1,273	62.25	7.13	Age at the time of the survey (in years)
Female	1,273	0.43	0.5	Equal to 1 for a female, 0 for a male
Married	1,273	0.86	0.35	Equal to 1 if married, 0 otherwise
Widowed	1,273	0.06	0.24	Equal to 1 if widowed, 0 otherwise
Single person	1,273	0.04	0.19	Not married, widowed or divorced; Equal to 1 if single, 0 otherwise
Divorced	1,273	0.04	0.2	Equal to 1 if divorced, 0 otherwise
Low education	1,273	0.28	0.45	At most a lower vocational training; Equal to 1 if low educated, 0 otherwise.
Medium education	1,273	0.38	0.49	Not low or high education; Equal to 1 if medium educated, 0 otherwise
High education	1,273	0.34	0.47	Higher vocational or university degree; Equal to 1 if high educated, 0 otherwise
Employed	1,273	0.24	0.43	Dummy variable (0-1).
Not employed	1,273	0.34	0.48	E.g. unemployed; dummy variable (0-1)
Retired	1,273	0.37	0.48	Dummy variable (0-1)
On disability	1,273	0.05	0.21	Dummy variable (0-1)
Home owner	1,273	0.72	0.45	Equal to 1 if home owner, 0 otherwise
Low household income	1,273	0.26	0.44	Equal to 1 if income is in the first tercile of the (population) distribution, 0 otherwise
Medium household income	1,273	0.26	0.44	Equal to 1 if income in second tercile, 0 otherwise
High household income	1,273	0.48	0.50	Equal to 1 if income in third tercile, 0 otherwise
Chronic illness	1,273	0.37	0.48	Equal to 1 if has a chronic illness, 0 otherwise
Smoking	1,273	0.25	0.43	Equal to 1 if a smoker, 0 otherwise
Alcohol	1,273	0.08	0.27	Equal to 1 if drinks four or more glasses a day
Bad health	1,273	0.05	0.22	Equal to 1 if bad health (dummy variable)
Overweight	1,273	0.41	0.49	25≤BMI<30; Dummy variable.
Obese	1,273	0.06	0.24	BMI≥30; Dummy variable.
Unhappy	1,273	0.15	0.36	Self-reported; Equal to 1 if unhappy, 0 otherwise
Mortality	1,273	0.49	0.50	Equal to 1 if died before 31/12/2018, 0 otherwise
Year of death	629	2009	6.35	Year of death, if mortality=1
SSP75	836	0.71	0.20	SPP until age 75; between 0 and 1
SSP80	1,060	0.56	0.22	SSP until age 80; between 0 and 1
SSP85	355	0.46	0.22	SSP until age 85; between 0 and 1
SSP90	195	0.32	0.23	SSP until age 90; between 0 and 1
SSP95	82	0.25	0.22	SSP until age 95; between 0 and 1
SSP100	18	0.15	0.22	SSP until age 100; between 0 and 1
ALE	1,273	84.04	2.05	Actuarial life expectancy (by age and gender)
DK-PLE	1,273	0.27	0.45	Equal to 1 if does not know PLE, 0 otherwise
PLE	926	77.44	3.72	For those with DK-PLE=0, believed PLE
PLE-knowledge	926	-6.5	3.53	PLE-ALE
ALE if DK-PLE=0	926	83.93	2.03	ALE if DK-PLE=0

Notes: N = Number of observations. Household income is equivalized using the square root of household size and the terciles are based on the distribution of household income in the raw data. All health information is self-reported. SSP=subjective survival probability. ALE = age and gender specific population remaining life expectancy based on actuarial life tables. PLE = (believed) population life expectancy. Age, ALE and PLE are measured in full years.

Appendix B   Objective and Subjective Mortality Risk Models

We assumed that lifetimes can be modeled with a Gompertz distribution (e.g., Gompertz 1825; Olshansky and Carnes 1997; Perozek 2008). The respondent's covariates are denoted by $\mathbf{x}$ and a proportional hazard specification is used to relate these covariates and age t to the annual mortality rate (Cox 1972):

$$\theta(t|\mathbf{x}; \gamma, \boldsymbol{\beta}) = \exp(\gamma t + \mathbf{x}\boldsymbol{\beta}). \tag{B1}$$

The parameters of interest are γ and $\boldsymbol{\beta}$, and $\mathbf{x}$ includes the covariates DK-PLE and PLE-knowledge. Age at the time of interview is denoted by t_0 and the probability of age at death being greater than t is given by the survival function

$$S(t|t_0, \mathbf{x}; \gamma, \boldsymbol{\beta}) = \exp\left(- \int_{t_0}^{t} \theta(s|\mathbf{x}; \gamma, \boldsymbol{\beta})ds\right). \tag{B2}$$

As age is observed in full years, the predicted remaining lifetime; conditional on t_0 and $\mathbf{x}$ is given by

$$L(t_0, \mathbf{x}; \gamma, \boldsymbol{\beta}) = \sum_{s=t_0}^{T} S(s|t_0, \mathbf{x}; \gamma, \boldsymbol{\beta}). \tag{B3}$$

T is the maximum age and set equal to 110 years. Given data on age of death, maximum likelihood estimates of γ and $\boldsymbol{\beta}$ are given by (Lancaster, 1990):

$$(\hat{\gamma}_O, \widehat{\boldsymbol{\beta}}_O) = \underset{\gamma, \boldsymbol{\beta}}{\mathrm{argmax}} \sum_{i=1}^{n} \ln\left(S(t_i|t_{0i}, \mathbf{x}_i; \gamma, \boldsymbol{\beta})\right) + (1 - c_i)\ln\left(\theta(t_i|\mathbf{x}_i; \gamma, \boldsymbol{\beta})\right), \tag{B4}$$

where i is an index for the individual and c is equal to one if the observation is right censored (i.e., the respondent died after December 31, 2018), and zero otherwise. In the main text this model is referred to as the objective mortality risk model.

Given subjective survival probabilities, estimates of γ and $\boldsymbol{\beta}$ were obtained using nonlinear least squares:

$$(\hat{\gamma}_S, \widehat{\boldsymbol{\beta}}_S) = \underset{\gamma, \boldsymbol{\beta}}{\operatorname{argmin}} \sum_{i=1}^{n} \sum_{j=1}^{2} \left(\ln(SSP_{ij}) - \ln\left(S(t_{ij}|t_{0i}, \mathbf{x}_i; \gamma, \boldsymbol{\beta}) \right) \right)^2, \tag{B5}$$

where SSP_{ij} is the jth subjective survival probability of respondent i and t_{ij} is the target age. Each respondent provided two SSPs. In the main text this model is referred to as the subjective mortality risk model.

The estimates from the objective and subjective mortality risk models, i.e. $(\hat{\gamma}_O, \widehat{\boldsymbol{\beta}}_O)$ and $(\hat{\gamma}_S, \widehat{\boldsymbol{\beta}}_S)$, respectively, were used to compute predicted (remaining) lifetimes (B3), conditional on the covariates $\mathbf{x}$ and age at interview t_0. The prediction of interest in the main analysis is the difference between the subjective and the objective predicted lifetimes, i.e. $L(t_0, \mathbf{x}; \hat{\gamma}_S, \widehat{\boldsymbol{\beta}}_S) - L(t_0, \mathbf{x}; \hat{\gamma}_O, \widehat{\boldsymbol{\beta}}_O)$ and how this difference was affected by knowledge on population life expectancy (one of the covariates). When computing standard errors, the covariances between $(\hat{\gamma}_O, \widehat{\boldsymbol{\beta}}_O)$ and $(\hat{\gamma}_S, \widehat{\boldsymbol{\beta}}_S)$ were taken into account (White 1980) and standard errors were clustered at a household level (Froot 1989).

Appendix C   Full Set of Estimation Results

Cells: Marginal changes (percentage changes in mortality risk)	Objective mortality model		Subjective mortality model	
Covariates	PE	SE	PE	SE
Female (relative to male)	-42.65***	(7.70)	9.88***	(2.14)
Age (one-year increase)	12.68***	(0.68)	7.23***	(0.23)
A one-year higher age at interview	4.71	(10.11)	0.11	(2.10)
Knowledge of population life expectancy (PLE)				
Does not know PLE (relative to knowing it)	6.51	(13.90)	20.90***	(3.42)
Difference between believed and actuarial PLE (a one-year increase)	-0.23	(1.38)	-3.03***	(0.27)
Socioeconomic status (dummy variables)				
Low educated	7.26	(11.11)	1.07	(2.03)
High educated	4.72	(11.89)	0.30	(2.11)
Not employed	8.38	(20.21)	1.00	(3.75)
Retired	14.74	(18.30)	18.25***	(4.43)
On disability	49.01	(30.85)	4.37	(6.01)
Home owner	-7.87	(9.05)	1.71	(1.92)
Low household income	6.42	(14.27)	-12.04***	(2.39)
High household income	-9.64	(10.00)	1.74	(2.19)
Marital status (dummy variables)				
Divorced	-6.38	(23.31)	-26.22***	(3.86)
Widowed	-6.05	(13.84)	-14.71***	(2.43)
A single person	-0.84	(22.09)	16.86***	(3.50)
Health/health behavior (dummy variables)				
A smoker	89.23***	(17.05)	14.99***	(2.31)
Drinks alcohol	46.11*	(19.80)	23.60***	(3.63)
Overweight	7.60	(9.87)	0.08	(1.77)
Obese	49.50†	(26.32)	-11.64***	(2.61)
Has a chronic illness	34.73**	(12.49)	20.28***	(2.42)
In bad health	33.97	(23.13)	73.68***	(3.52)
Feels unhappy	31.78*	(15.63)	17.29***	(2.38)
Mean log-likelihood	-2523		-1608	
Number of individuals	1,273		1,273	
Test results	p-value		p-value	
H_0: No relationship with knowledge on PLE	0.863		0.000***	
H_0: No relationship with socioeconomic status	0.285		0.000***	
H_0: No relationship with marital status	0.965		0.000***	
H_0: No relationship with health (behavior)	0.000***		0.000***	

Notes: A marginal change is the percentages difference in the annual mortality rate compared to a reference individual. The reference individual is a 55-year-old male at the time of interview, medium educated, employed, and married, with medium household income, who did not smoke or drink, had a normal body weight, no chronic illnesses, felt happy, and reported to be in good health. PE = Parameter estimate; SE = Standard Error. Levels of significance: †p ‹ 0.10; *p ‹ 0.05; **p ‹0.01; ***p ‹0.001.

FIGURES AND TABLES

**Fig. 1** Observed and average subjective survival probabilities by target age.
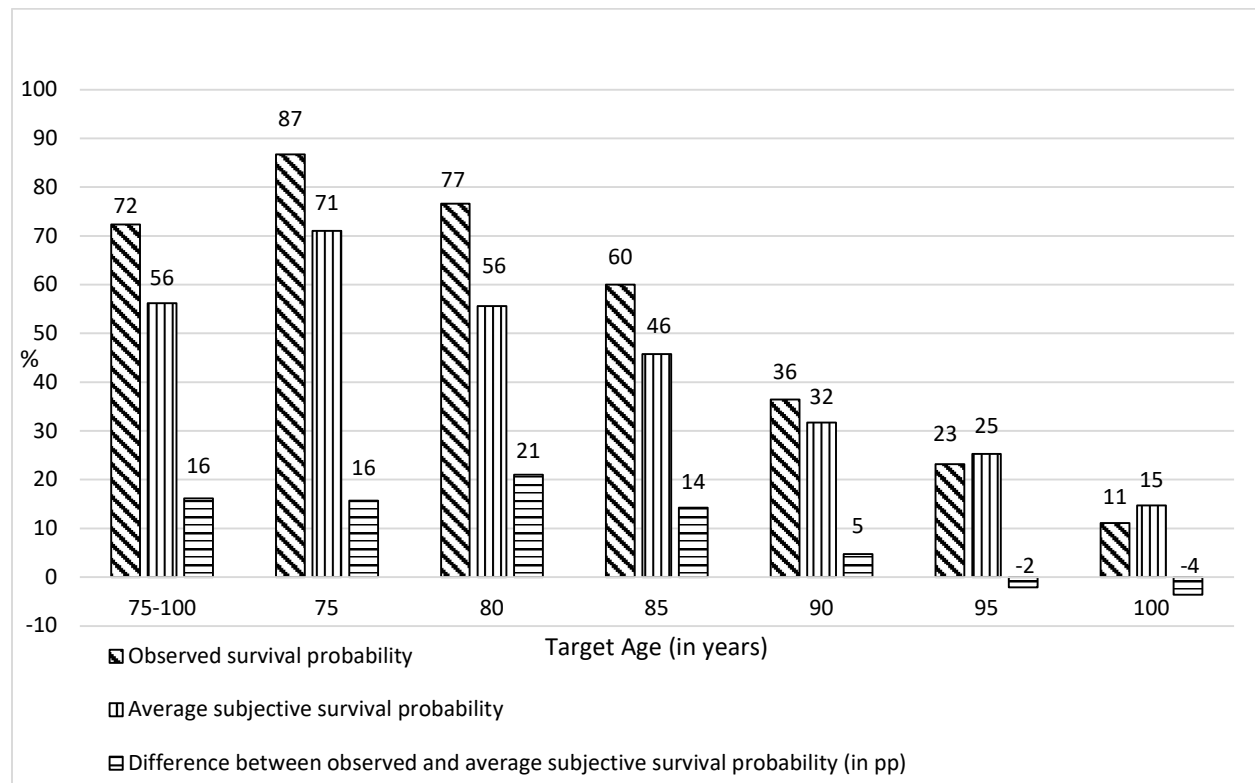
Notes: The observed survival probability is the proportion alive at a specific target age of those who reported SSPs for that target age (in percentages). The average survival probability is the average SPPs reported by those individuals (in percentages). At the top of the bars are percentages or percentage points (pp). The number of observations for target ages 75, 80, 85, 90, 95 and 100 are, respectively, 836, 1,060, 355, 195, 82 and 18.

Fig. 2 Respondents' beliefs and actuarial predictions of population life expectancy.
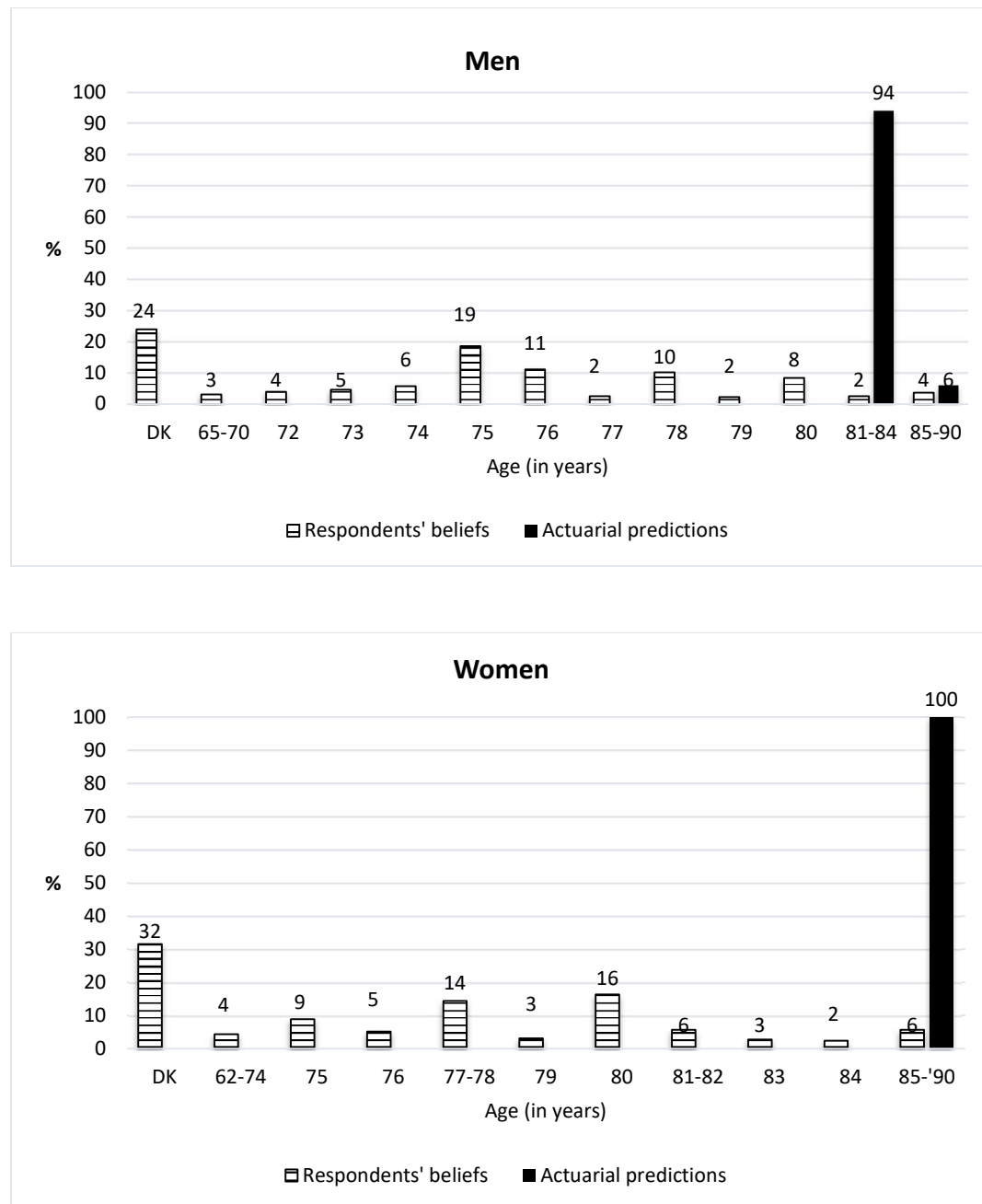
Notes: DK = Don't know remaining population life expectancy (DK-PLE=1). As the survey questions refer to remaining life expectancy, these numbers are conditional on age at time of the survey. At the top of the bars are percentages.

Fig. 3 The variable PLE-knowledge: Respondents' beliefs minus actuarial predictions of remaining population life expectancy (reported on in Fig. 2).
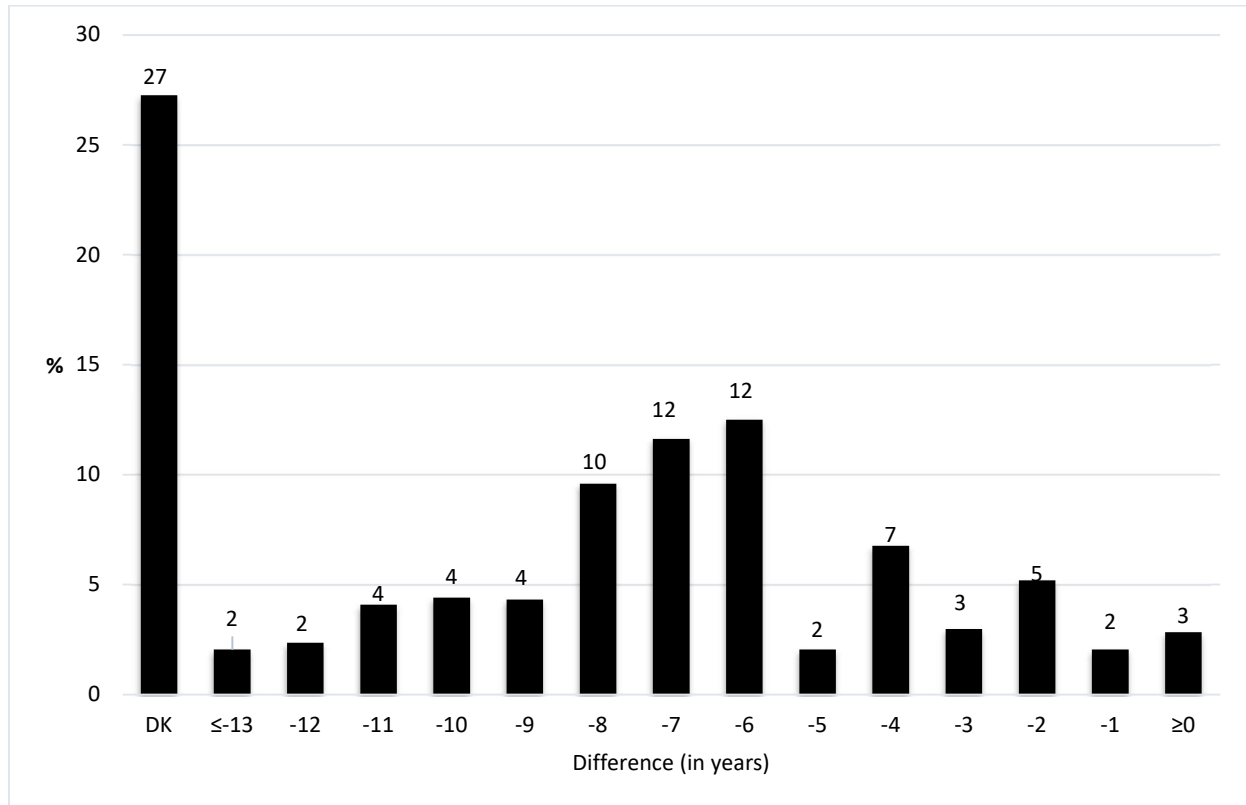
Note: DK = Don't know remaining population life expectancy (DK-PLE=1). At the top of the bars are percentages.

Fig. 4 Respondents' beliefs and observed survival probabilities by degree of knowledge of population life expectancy.
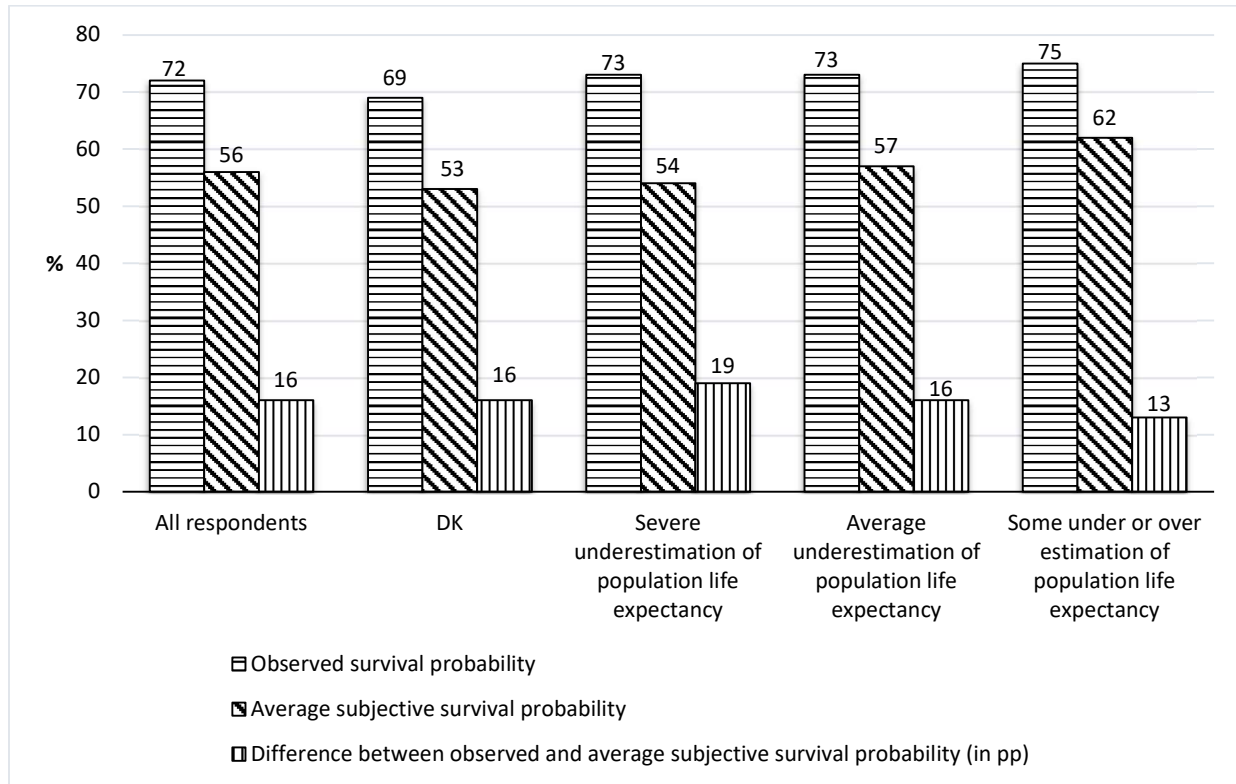
Notes: DK = Don't know remaining population life expectancy (DK-PLE=1). Severe underestimation of population life expectancy is defined, for those with DK-PLE=0, as PLE-knowledge<-8, average underestimation as -8 ≤ PLE-knowledge ≤ -6, and some under- or overestimation as PLE-knowledge ≥ -5. Only about 2% of respondents overestimated population life expectancy and this group is too small for considering it separately. At the top of the bars are percentages or percentage points (pp).

Table 1   Summary of the estimation results: Predicted mortality rates by age and the marginal effects of various covariates on the objective and subjective mortality rates.

	Objective mortality model		Subjective mortality model	
Panel A: Predicted annual mortality rates				
(In percentages and for selected ages)[a]	PE	(SE)	PE	(SE)
Age 55	0.13**	(0.03)	0.89**	(0.06)
Age 65	0.42**	(0.10)	1.83**	(0.09)
Age 75	1.52**	(0.40)	3.78**	(0.21)
Age 85	5.47**	(1.67)	7.79**	(0.53)
Age 95	19.79**	(7.13)	16.08**	(1.38)
Panel B: Marginal changes				
(percentage differences in the mortality rate)	PE	(SE)	PE	(SE)
Female (relative to male)	-42.65**	(7.70)	9.88**	(2.14)
Age gradient (a one-year increase)	12.68**	(0.68)	7.23**	(0.23)
Knowledge of population life expectancy (PLE)				
Does not know it (relative to knowing it exactly)	6.51	(13.9)	20.90**	(3.42)
A one-year increase in the difference between believed and actuarial population life expectancy	0.23	(1.38)	3.03**	(0.27)
Number of individuals	1,273		1,273	
Panel C: Test results	p-value		p-value	
H_0: No relationship with knowledge of PLE	0.86		0.00**	
H_0: No relationship with socioeconomic status	0.28		0.00**	
H_0: No relationship with marital status	0.97		0.00**	
H_0: No relationship with health (behavior)	0.00**		0.00**	

Notes: PE = Parameter estimate; SE = Standard Error. Levels of significance: †p ‹ 0.10; *p ‹ 0.05; **p ‹0.01; ***p ‹0.001

[a] These predictions are for a reference individual who is male, medium educated, employed, married, and with medium household income, who did not smoke or drink, reported to be in good health, felt happy, and had a normal body weight and no chronic illnesses. All covariates affected the annual mortality rate proportionally.

Table 2   Predicted lifetimes and marginal changes in the accuracy of predicted subjective lifetime due to changes in the socioeconomic status and health-related covariates.

	PE	SE
Panel A: predictions for a reference individual [a]		
Subjective lifetime (*SL*), in years	81.12***	(0.46)
Objective lifetime (*OL*), in years	87.85***	(1.52)
Subjective minus objective lifetime (SL-OL), in years	-6.73***	(1.62)
Panel B: Marginal changes in the difference (SL-OL), in years		
Female	-5.23***	(1.04)
Age at interview (a one-year increase)	0.03	(0.08)
Low educated	0.41	(0.84)
High educated	0.32	(0.92)
Not employed	0.48	(1.48)
Retired	-0.74	(1.28)
On disability	2.52	(1.74)
Homeowner	-0.81	(0.80)
Low household income	1.85†	(1.09)
High household income	-0.96	(0.90)
Divorced	2.80	(2.16)
Widowed	1.25	(1.24)
Single person	-1.70	(1.71)
A smoker	3.29***	(0.76)
Drinks alcohol	0.63	(1.08)
Overweight	0.54	(0.74)
Obese	4.34**	(1.42)
Has a chronic illness	0.31	(0.74)
In bad health	-3.42*	(1.35)
Unhappy	0.41	(0.93)

Notes: PLE = Population life expectancy. All covariates except for age are dummy variables. PE = Parameter estimate; SE = Standard Error. Levels of significance: †$p < 0.10$; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

[a] The reference individual is a 55-year-old male, medium educated, employed, and married, had medium household income, did not smoke or drink alcohol, had a normal body weight, and had no chronic illnesses, felt happy, and reported to be in good health.

Table 3 **Marginal changes in the accuracy of predicted subjective lifetime due to**

changes in knowledge of population life expectancy (PLE) and robustness checks.

Marginal changes (in years)	PE	SE
Main specification I: Tables 1 & 2		
Does not know PLE	0.30	(0.78)
Knows PLE exactly	1.87**	(0.68)
Knows PLE & underestimates it with 6 years (reference individual)	0.00	(-)
Knows PLE & underestimates it with 5 years	0.31**	(0.11)
Specification II: Mortality models without controlling for socioeconomic and health variables		
Does not know PLE	1.08	(0.72)
Knows PLE exactly	1.77*	(0.71)
Knows PLE & underestimates it with 6 years (reference individual)	0.00	(-)
Knows PLE & underestimates it with 5 years	0.29*	(0.12)
Specification III: PLE-knowledge modelled as a spline function (knots at -12 and 0 years)		
Does not know PLE	0.45	(0.81)
Knows PLE exactly	3.33**	(1.24)
Knows PLE & underestimates it with 6 years (reference individual)	0.00	(-)
Knows PLE & underestimates it with 5 years	0.35*	(0.14)
Specification IV: Possible outliers, removed 56 observations with *PLE-knowledge* <-12 or >0		
Does not know PLE	0.49	(0.79)
Knows PLE exactly	2.77**	(0.89)
Knows PLE & underestimates it with 6 years (reference individual)	0.00	(-)
Knows PLE & underestimates it with 5 years	0.45**	(0.15)
Specification V: dropped 347 individuals who did not know PLE (DK-PLE=1)		
Does not know PLE		
Knows PLE exactly	1.99**	(0.65)
Knows PLE & underestimates it with 6 years (reference individual)	0.00	(-)
Knows PLE & underestimates it with 5 years	0.33**	(0.11)
Specification VI: Only women (551 observations; 228 died)		
Does not know PLE	-0.04	(1.23)
Knows PLE exactly	0.86	(1.25)
Knows PLE & underestimates it with 6 years (reference individual)	0.00	(-)
Knows PLE & underestimates it with 5 years	0.14	(0.21)
Specification VII: Only men (722 observations; 401 died)		
Does not know PLE	0.38	(1.05)
Knows PLE exactly	2.97**	(0.92)
Knows PLE & underestimates it with 6 years (reference individual)	0.00	(-)
Knows PLE & underestimates it with 5 years	0.49**	(0.15)

Notes: PLE=Population Life Expectancy; PLE-knowledge=believed minus actuarial PLE. PE = Parameter estimate;

SE = Standard Error. Levels of significance: †$p < 0.10$; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$