Test-Retest Reliability of Subjective Survival Expectations*

Jochem de Bresser
Tilburg University†
Netspar‡

October 8, 2016

Abstract

This paper analyzes the test-retest reliability of subjective survival expectations. Using a nationally representative sample from the Netherlands, we compare probabilities reported by the same individuals in two different surveys that were fielded in the same month. We evaluate reliability both at the level of reported probabilities and through a model that relates expectations to socio-demographic variables. Test-retest correlations of survival probabilities are between 0.5 and 0.7, which is similar to subjective well-being (Krueger and Skade, 2008). Only 20% of probabilities are equal across surveys, but up to 61-77% are consistent once we account for rounding. Models that analyze all probabilities jointly reveal that similar associations emerge between covariates and the hazard of death in both datasets. Moreover, expectations are persistent at the level of the individual and this unobserved heterogeneity is strongly correlated across surveys. Finally, we use a life-cycle model to map survival expectations into simulated wealth and labor supply. Though wealth accumulation is sensitive to survival expectations, simulated probabilities from a model that corrects for rounding are sufficiently reliable to yield reliable wealth profiles. Taken together this evidence supports the reliability of subjective survival expectations.

Key words: Subjective expectations, test-retest reliability, life-cycle model, rounding

JEL-codes: D84; J14; C34

*This work is part of the research programme Innovational Research Incentives Scheme Veni with project number 451-15-018, which is (partly) financed by the Netherlands Organisation for Scientific Research (NWO). I also thank Netspar for financially supporting data collection. All errors are my own.

†Tilburg University, P.O. Box 90153, 5000 LE Tilburg the Netherlands. Email: j.r.debresser@uvt.nl.
‡Netspar, P.O. Box 90153, 5000 LE Tilburg, The Netherlands

1 Introduction

Expectations play an important role in economic models of inter-temporal decision making, such as life-cycle models of labor supply and saving (e.g. French, 2005; De Nardi et al., 2010; French and Jones, 2011). Over the past two decades researchers have started to recognize the potential of data that measure subjective expectations held by survey respondents, especially when elicited in terms of probabilities (see Manski, 2004, for a review). However, the validity of such intrinsically subjective data remains controversial. This paper is the first to evaluate the test-retest reliability of expectations reported by survey respondents and to link this reliability to economic behavior through a structural model. We focus on expectations regarding one's own survival and compare the responses of the same individuals in the same month between two surveys, both of which measure a number of points on the subjective survival curve.

Our data have been collected in the CentERpanel, a large household panel that is representative for the Dutch population. The two surveys analyzed are the Pension Barometer (PB) and the DNB Household Survey (DHS). Given that each survey elicits beliefs by means of multiple survey items, reliability can be gauged at two levels. Firstly, we check whether the probabilities are consistent with each other one-by-one. We compare probabilities reported by the same individuals for the same target ages, taking into account differences between answer scales. Secondly, we formulate a model in which we use all reported probabilities simultaneously to look at the relationships between subjective survival and background variables. We assess to what extent the two sets of probabilities yield similar associations between the hazard of death and socio-economic covariates when analyzed jointly.

Having quantified the reliability of subjective survival expectations, we evaluate whether they are sufficiently reliable to be used as inputs for life-cycle models of saving and labor supply. A life-cycle model maps survival curves into behavior. We check whether such simulated behavior is sensitive to variation in survival curves of the magnitude observed between surveys.

This paper builds a bridge between the literature on subjective expectations, as summarized by Hurd (2009), and that on empirical life-cycle models. A rich body of research has established the covariates and predictive validity of survival expectations at the level of the individual (Hurd and McGarry, 1995, 2002; Smith et al., 2001; Bissonnette et al., 2011; Kutlu and Kalwij, 2012). To date, plausible associations between subjective survival and background variables provide the most important evidence in support of the validity of this type of data. However, the way questions are framed does affect reported expectations: a "die by" frame yields lower life expectancy than does a "live to" frame (Payne et al., 2013; Teppa et al., 2015). Survival expectations play an important role in the saving decisions analyzed in structural life-cycle models. Researchers typically equate those expectations to actuarial forecasts, despite the robust finding that such figures are a poor proxy for average expectations (Bissonnette et al., 2011; Perozek, 2008).

Our analysis contributes to the literature in different ways. We assess the validity of subjective survival in a nationally representative panel, while earlier studies have tended to focus on older cohorts (the HRS and SHARE are often used yet they only cover the 50-plus population). Moreover, test-retest analysis has been applied to survey data of various types, such as wellbeing (Krueger and Skade, 2008). Hence, it allows one to compare the reliability of elicited beliefs to that of other, more commonly used types of data. We analyze reliability at different levels of aggregation: multiple probabilities for each individual-year allow us to investigate whether discrepancies between reported probabilities cancel out when probabilities are combined to fit survival curves. Furthermore, we observe multiple probabilities reported by an individual in a survey in addition to repeated observations for the same individual. These two levels of clustering allow us to disentangle the reliability of variation in beliefs for a given individual over time (within-variation) from the reliability of variation across individuals (between-variation). We take into account the specific measurement error that comes from rounding, either survey-induced or not, in a comprehensive way. Finally, we simulate saving and labor supply to give economic meaning to our analysis of reliability.

We find that reported probabilities are reliable overall. Our analysis of individual probabilities yields test-retest correlations between 0.5 and 0.7, which is comparable to the reliability of subjective well-being documented by Krueger and Skade (2008). While only around 20% of reported probabilities are exactly equal, 25-37% are consistent when we account for the different resolutions of response scales. Rounding further increases the rate of consistent responses to 32-46% if we assume all probabilities reported by a given respondent are rounded similarly and 61-77% if we allow for the maximum degree of rounding for each reported probability. Models in which all reported probabilities are analyzed jointly show that the associations between the hazard of death and most socio-demographic covariates are similar for both datasets. However, substantially different associations are found for the covariate *birth cohort*, especially for older cohorts. Individual effects account for 90% of variation that cannot be explained by demographic covariates and are strongly correlated between surveys (correlation coefficients 0.8-0.9). The correlation between survey-effects that account for the remaining 10% is much lower, suggesting that the variation in beliefs across individuals is more reliable than longitudinal variation for a given individual. Accounting for rounding improves model fit. Unlike labor supply, the level of wealth accumulated in a life-cycle model is sensitive to survival expectations. However, using the model that accounts for rounding we obtain sufficiently similar survival curves to simulate wealth reliably. Hence, we conclude that reported expectations are sufficiently reliable to be used in structural models.

The rest of the paper is structured as follows. Section 2 describes our data in detail and section 3 evaluates the reliability of the reported probabilities one by one. Section 4 presents the model used to analyze all probabilities jointly, after which section 5 presents estimation results. We evaluate the economic significance of differences between the two datasets by means of simulated life-cycle models in section 6, after which section 7 concludes.

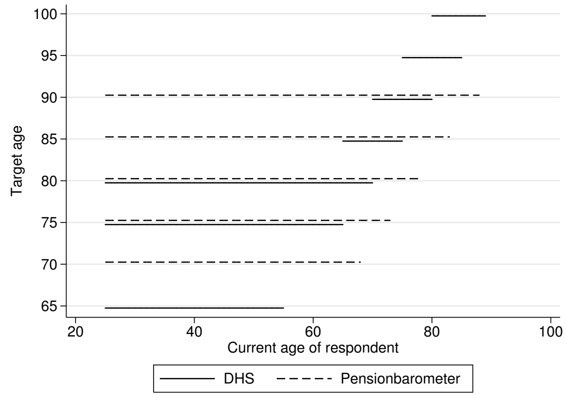

Figure 1: Age eligibility for survival questions in the DHS and in the Pensionbarometer

2 Survival questions in the Pension Barometer and in the DNB Household Survey

Both the PB and the DHS were administered to the CentERpanel. The CentERpanel is a household panel that is representative for the Dutch population and that is managed by CentERdata at Tilburg University. In both surveys respondents are offered multiple survival questions asking for the likelihood of surviving to different target ages based on their current age. Figure 1 shows graphically which ages are eligible for each question in both questionnaires. As can be seen in that figure, the PB elicits expectations for five equally spaced target ages between 70 and 90, while the DHS asks questions about age 65 and six ages between 75 and 100. Hence, we can directly compare probabilities corresponding to the target ages 75, 80, 85 and 90. The PB offers survival questions to respondents of age 25 and older who are at least 2 years younger than the target age for which expectations are elicited. Hence, the potential sample for the PB is larger for questions referring to older ages and respondents of age 68 and younger are offered all five survival questions included in the survey. The DHS, on the other hand, asks one or two questions according to the age of the respondent.

Other than the response format, questions are phrased similarly in the PB and the DHS. The PB asks:

> "Please indicate on a scale from 0 to 100 how likely you think it is that you
> *[If age < 69]* will live to age 70."
> etc.

The items in the DHS are phrased as follows:

> "Please indicate your answer on a scale of 0 thru 10, where 0 means 'no chance at all' and 10 means 'absolutely certain'.
> How likely is it that you will attain (at least) the age of 65?"
> etc.

In the PB the questions are preceded only by a single item on subjective health, asking respondents to rate their health on a 5-point scale from 'excellent' to 'poor'. The DHS questionnaire contains 14 questions before the survival questions, which are the final questions to be asked in the health-section of the survey. In addition to a question on subjective health that is identical to that in the PB, the DHS also includes questions on height, weight, consumption of alcohol and cigarettes, doctor visits and absenteeism due to health problems.

3 Reliability of reported probabilities

3.1 Descriptives

Before setting up a formal model, we investigate the extent to which the reported probabilities are consistent with each other for the same individuals and target ages. For most individuals both surveys were conducted in June of 2011 and 2012. The notion that both questionnaires aim to measure the same expectations is plausible, since the period between questionnaires is

Table 1: Descriptive statistics of the reported survival probabilities and life table (LT) probabilities

	N	Current age	Mean LT	PB Mean	PB S. D.	DHS Mean	DHS S. D.	Rank corr.
a. Men								
Age 75	823	25-63	75.2	65.3	23.0	68.0	19.2	0.66
Age 80	1000	25-68	60.6	52.7	24.9	55.7	22.7	0.68
Age 85	294	65-73	45.7	40.9	25.8	52.5	22.9	0.58
Age 90	188	70-78	25.1	26.4	24.6	38.5	24.6	0.55
b. Women								
Age 75	690	25-63	83.6	65.8	22.5	67.5	19.0	0.56
Age 80	796	25-68	73.7	55.1	24.7	57.0	22.0	0.56
Age 85	168	65-73	61.7	44.5	26.0	54.0	23.0	0.61
Age 90	103	70-78	40.0	29.7	25.0	39.5	24.3	0.53

short. In 2,187 matched individual-year records the average time between surveys is 3.3 weeks with a median of 1 week and no more than 4 weeks between questionnaires for over three quarters of observations. Both surveys took place in the same week for 6% of person-year observations.[1]

Rates of non-response and logically consistent answers are similar across the two surveys. 95% of age-eligible respondents answer all relevant PB survival questions compared with 91% for the DHS. Moreover, 98% of the responses to the PB questions and 99% of responses to DHS questions decrease weakly with age and are thus logically consistent. Out of 2,988 potential observations for the PB, we are left with 2,781 complete and consistent person/year observations. Similarly, 3,584 observations for the DHS yield 3,246 useful observations. In the remainder of this section we limit ourselves to the 2,187 observations for which we observe complete and monotonic response to both the PB and the DHS. Due to different age-eligibility rules for the various target ages in the questionnaires, we have 2,087 observations for which we observe at least one reported probability for the same target age.

[1]In the paper we report results using all records that could be matched, regardless of the time between surveys. Robustness checks indicate that none of our findings change when we limit the sample to cases for which the two surveys were taken within a 4-week period.

Table 1 shows descriptives of reported subjective probabilities and corresponding probabilities from the 2010 life tables published by Statistics Netherlands.[2] Summary statistics are presented by target age and for each target age we limit the sample to those respondent-years that reported a probability in both surveys. Looking first at the means of the probabilities reported in the PB and in the DHS, we observe that the means are close together for the target ages of 75 and 80 (differences are less than 3 percentage points). However, for the older target ages the average probability in the DHS is around 10 percentage points higher than that in the PB. As a result the average DHS probability is higher than the life-table forecast for ages 85 and 90 for men. Women report probabilities that are substantially below actuarial predictions for all ages, so for them the DHS yields expectations that are more in line with official forecasts. The (rank) correlations between PB and DHS probabilities are between 0.53 and 0.68, which is similar to that found for subjective well-being (Krueger and Skade, 2008). Hence, based on the correlations between reported probabilities the reliability of subjective survival expectations is comparable to that of another widely researched type of subjective data, even though the levels are different for older target ages. Note, however, that while a given aspect of well-being is usually measured by a single item in a questionnaire, there is scope to combine the various reported probabilities and construct survival functions.

Figure 2 shows the medians and inter-quartile ranges of the distributions of PB probabilities conditional on a certain response to the DHS items by target age. The figures confirm that both sets of probabilities are closely related for the target ages 75 and 80: medians are mostly close to the diagonal and IQRs are relatively narrow (around 20 %-points). For the target ages of 85 and 90 the correspondence between the two is less tight, especially among those respondents who indicate a relatively large chance of 40% or higher of surviving past those ages in the DHS. The medians of the distributions of PB probabilities are 10-30 per-

[2]The life-tables are matched based on gender and age at the time of the survey, so differences between the age distribution of the Dutch population and that of the subsample that answers a particular question do not affect the comparison.

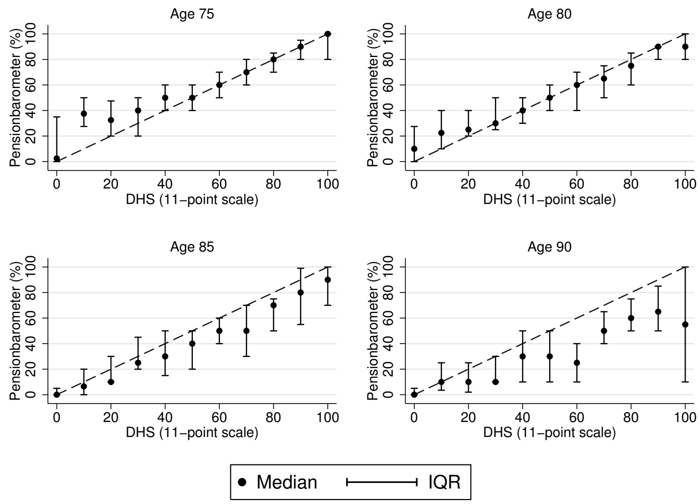

Figure 2: Medians and IQRs of survival probabilities in the Pension Barometer conditional on responses to the corresponding DHS question

centage points below the diagonal and even the third quartile is often below the diagonal, indicating that more than 75% of respondents who are relatively certain to survive past 85 or 90 according to the DHS report less certainty in the PB.

3.2 One-by-one reliability

The most intuitive way to compare PB and DHS probabilities may be to look at the distribution of the differences between the two. However, the possibility of rounding implies that the (absolute) difference between reported probabilities is not a good measure of the extent to which the data are compatible. For instance, reported probabilities of 100% in the DHS and 55% in the PB are consistent if the former is rounded to a multiple of 100 (so that the true probability lies in $[100, 50]$). On the other hand, probabilities of 65% and 55% would be incompatible, since both are only consistent with rounding to multiples of 1 or 5 and thus the intervals for the true probability do not overlap.

Therefore, our approach is to determine the extent of rounding based on three different rounding schemes and to check whether the probabilities reported in the PB and the

Table 2: Rates of consistent responses to PB and DHS survival questions

	N	Exactly equal	Minimal rounding	Common rounding	General rounding
Age 75	1513	0.22	0.37	0.46	0.77
Age 80	1796	0.22	0.31	0.40	0.75
Age 85	462	0.18	0.26	0.34	0.68
Age 90	291	0.16	0.24	0.32	0.61
All combined[a]	2,087	0.09	0.18	0.27	0.63

[a] The sample size is 2,087 individual-years rather than 2,187 as mentioned above, since we exclude observations for which we have monotonic and complete probabilities for both the PB and the DHS, but for which the two questionnaires have no target ages in common.

DHS can reflect the same underlying true probability under each of those rules. The first scheme assumes that each probability is reported as precisely as allowed by each survey: all probabilities in the PB are rounded to multiples of 1 and all probabilities in the DHS to multiples of 10. Hence, under this *minimal* rounding rule any two probabilities are compatible if $P^{PB} \in \left[P^{DHS} - 5, P^{DHS} + 5\right]$.[3] The second, *common*, scheme allows for more rounding, but maintains that all survival probabilities reported by the same individual are rounded similarly. We distinguish between the levels of rounding proposed by Manski and Molinari (2010) and refer the reader to that paper for more information. Finally, the third *general* rounding rule allows each reported probability to be rounded to the maximum extent (see Bissonnette and de Bresser, 2014, for more information on this scheme). Table A1 in Appendix A shows the distribution of rounding in the sample according to both rounding rules. Under common rounding we find that rounding to multiples of 5 is the most prevalent type for the PB, while rounding to multiples of 10 is most prevalent for the DHS (58% of individual-year observations of the PB are rounded to multiples of 5, while 95% of DHS observations are rounded to multiples of 10). For general rounding at the level of the individual probability, rounding to multiples of 10 is the most frequent category (52% of PB probabilities and 76% of DHS probabilities are rounded to multiples of 10).

[3] $P^{PB} = 15$ is consistent with $P^{DHS} = 10$ *and* $P^{DHS} = 20$, since the true PB probability may be anywhere in $[14.5, 15.5)$.

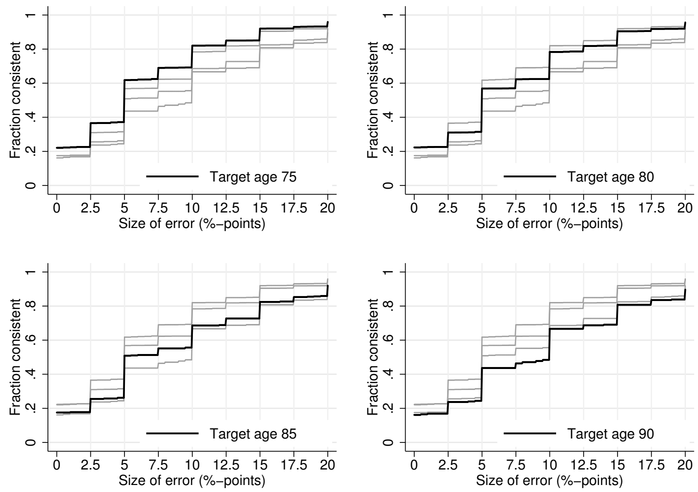

Figure 3: Fraction of probabilities that are consistent across PB and DHS while allowing for reporting noise

The rates of compatible responses to PB and DHS questions by target age and for the different rounding rules are given in Table 2. Around one fifth of reported probabilities are equal across surveys. If we assume that all probabilities are rounded to the minimal extent allowed by each survey we find a rate of consistent response that declines more steeply for older target ages from 37% for target age 75 to 24% for age 90. Allowing for common rounding increases the rate of consistent probabilities to 32-46%. Under the most conservative general rounding scheme 61-78% of responses are compatible with at least one underlying true probability. Regardless of the rounding rule, we find that the fraction of consistent responses is higher for younger target ages. These differences are mostly related to the current age of the respondents, rather than the target age to which questions refer. The rate of consistent answers to the two sets of questions is flat up to age 68 and declines sharply afterwards. Interestingly, the rate of consistent probabilities is the same when we restrict the sample to those observations that report the same level of subjective health in both survey waves or to surveys taken within a four week period. Hence, differences probably reflect measurement error rather than changes in the actual expectations held by respondents.

The upshot of the comparison so far is that while the two sets of probabilities are fairly strongly correlated, it takes considerable rounding error for a majority of the cases in order to make the PB and DHS responses compatible with at least one underlying true probability. Figure 3 illustrates this point is a slightly different way, showing how the fraction of reported probabilities that is consistent between the PB and the DHS increases with the size of a symmetric reporting error added to both probabilities. It takes a reporting error of 5 percentage points around the reported PB and DHS probability to make more than 40% of the pairs of probabilities compatible, while it takes an error of 10 percentage points to make 70-80% compatible. Note that even for an error of 20 percentage points over 15% of reported probabilities for the target ages 85 and 90 are irreconcilable.

These differences between the two sets of probabilities when analyzed one by one raise the question whether an analysis of all probabilities jointly would yield different results when based on the PB versus the DHS. In the next section we set up two models to answer that question.

4 Reliability of survival curves

4.1 Model without rounding

The model we use in this paper is closely related to that proposed by Kleinjans and Van Soest (2014) for expectations regarding binary outcomes and extended to continuous outcomes in De Bresser and Van Soest (2013). We refer the reader to those papers for more elaborate descriptions.

Expectations follow a Gompertz distribution with the baseline hazard shifted proportionally by demographic variables. We model expectations over complete lifespans and take into account truncation at the current age of the respondent. This parameterization of expectations implies that true probabilities of surviving to target age ta_k conditional on having survived to current age a_{it} are given by:

$$S^q_{itk}|a_{it} = \Pr\left(t \geq ta_k | t \geq a_{it}\right) = \frac{\Pr\left(t \geq ta_k \ \& \ t \geq a_{it}\right)}{\Pr\left(t \geq a_{it}\right)} = \frac{\Pr\left(t \geq a_{it} | t \geq ta_k\right) \times \Pr\left(t \geq ta_k\right)}{\Pr\left(t \geq a_{it}\right)}$$
$$= \frac{1 \times \Pr\left(t \geq ta_k\right)}{\Pr\left(t \geq a_{it}\right)} = \frac{\exp\left(-\frac{\gamma^q_{it}}{\alpha^q}\left(\exp\left(\alpha^q\left(ta_k/100\right)\right) - 1\right)\right)}{\exp\left(-\frac{\gamma^q_{it}}{\alpha^q}\left(\exp\left(\alpha^q\left(a_{it}/100\right)\right) - 1\right)\right)} \times 100$$

where q indexes questionnaires ($q \in \{PB, DHS\}$); $\gamma^q_{it} = \exp\left(\mathbf{x}'_{it}\boldsymbol{\beta}^q_1 + \xi^q_i + \eta^q_{it}\right)$ depends on the demographics of respondent i in survey-year t; α^q determines the shape of the baseline hazard; ta_k is a target age in the questionnaire and a_{it} is the age of i in year t. We distinguish two types of unobserved heterogeneity: individual effects ξ^q_i and question sequence effects η^q_{it}. Distributional assumptions for these error components are given later. In the absence of unobserved heterogeneity the null hypothesis of interest is that $\boldsymbol{\beta}^{PB}_1 = \boldsymbol{\beta}^{DHS}_1$ and $\alpha^{PB} = \alpha^{DHS}$, which implies that the two surveys yield the same associations between covariates and survival. We divide both the target age and the current age by 100 to facilitate estimation of α^q (which determines the shape of the baseline hazard).

However, we do not observe S^q_{itk} directly. Instead, the reported probabilities are perturbed by recall error:

$$P^{*q}_{itk} = S^q_{itk} + \varepsilon^q_{itk}$$

where $\varepsilon^q_{itk} \sim \mathcal{N}\left(0, \sigma^2_{it}\right)$, independent of all covariates and across thresholds, surveys, years and individuals. We model the variance of recall errors as $\ln\left(\sigma_{it}\right) = \mathbf{x}'_{it}\boldsymbol{\beta}^q_2$. In the baseline model we do not allow for rounding in the reported probabilities, but we do take into account censoring between zero and the lowest probability reported previously in the sequence. Hence, the density for a reported probability P_{itk}^{q} conditional on covariates is given by

$$f\left(P_{itk}^{q}|\mathbf{x}_{it}\right)=\begin{cases}1-\Phi\left(\frac{P_{it,k-1}^{q}-S_{itk}^{q}}{\sigma_{it}}\right) & \text{if } P_{itk}^{q}=P_{it,k-1}^{q} \text{ (censored from above)}\\ \phi\left(\frac{P_{itk}^{q}-S_{itk}^{q}}{\sigma_{it}}\right) & \text{if } 0<P_{itk}^{q}<P_{it,k-1}^{q} \text{ (uncensored)}\\ \Phi\left(\frac{P_{itk}^{q}-S_{itk}^{q}}{\sigma_{it}}\right) & \text{if } P_{itk}^{q}=0 \text{ (censored from below)}\end{cases}$$

where $\phi\left(.\right)$ and $\Phi\left(.\right)$ respectively denote the standard normal density and CDF and for the first threshold $k=1$ we set $P_{it0}^{q}=100$ (when estimating the model we also condition on individual and survey effects, but we omit them here for ease of exposition).

The model is completed by distributions of the individual effects ξ_{i}^{q} and survey effects η_{it}^{q}. We assume that both are bivariate normal with covariance matrices Σ_{ξ} and Σ_{η} and that they are independent of covariates and each other. We estimate the elements of the covariance matrices of unobserved heterogeneity, the baseline hazards α^{PB} and α^{DHS} and the vectors $\boldsymbol{\beta}_{1}^{PB}$, $\boldsymbol{\beta}_{2}^{PB}$, $\boldsymbol{\beta}_{1}^{DHS}$ and $\boldsymbol{\beta}_{2}^{DHS}$ by maximum simulated likelihood where we integrate numerically over the distributions of individual and question sequence effects.

4.2 Model with rounding

The basic setup is the same as for the baseline model, but now P_{itk}^{*q} is not only censored but also rounded prior to being reported. We allow for rounding to multiples of 100, 50, 25, 10, 5 and 1 for the pensionbarometer and to multiples of 100, 50 and 10 for the DHS. Our rounding model is ordinal:

$$R_{itk}^{q}=r \iff \mu_{r-1}^{q}\leq y_{it}^{*q}=\mathbf{x}_{it}'\boldsymbol{\beta}_{3}^{q}+\xi_{i}^{r,q}+\eta_{it}^{r,q}+\varepsilon_{itk}^{r}<\mu_{r}^{q}$$

where $r\in\{1,2,...,6\}$ for the PB and $r\in\{1,2,3\}$ for the DHS, with 1 being the least amount of rounding allowed by the survey. The rounding equation includes individual and question sequence effects, allowing rounding to be correlated across repeated observations for a given individual and to be more strongly correlated within than between survey waves. Moreover, both types of unobserved heterogeneity may be correlated across surveys (PB and DHS) and with their respective counterparts in the equation that shifts survival curves (ξ_i^{PB}, ξ_i^{DHS}, $\xi_i^{r,PB}$ and $\xi_i^{r,DHS}$ follow a four dimensional normal distribution and so do the survey effects $\boldsymbol{\eta}_{it}$). We assume that the idiosyncratic rounding shocks ε_{itk}^r follow a standard normal distribution and are independent from covariates and all other errors, so the conditional probabilities of each category of rounding $\Pr\left(R_{itk}^q = r|\mathbf{x}_{it}, \boldsymbol{\xi}_i, \boldsymbol{\eta}_{it}\right)$ take the shape of an ordered probit.

A reported probability in combination with a particular level of rounding implies an interval for the perturbed probability $P_{itk}^{*q} \in [LB_{itk}^r, UB_{itk}^r)$. For instance, a reported probability of 25% that is rounded to a multiple of 5 yields the interval $P_{itk}^{*q} \in [22.5, 27.5)$. The probability of that event is easy to calculate, since $P_{itk}^{*q} \sim \mathcal{N}\left(S_{itk}^q, \sigma_{it}^2\right)$. As a given reported probability may result from different degrees of rounding, rounding is a latent construct and we average across the different degrees of rounding to obtain the likelihood contribution. In particular, define for each reported probability the set Ω_{itk} that consists of all types of rounding that are consistent with that probability. We obtain the conditional density as (omitting unobserved heterogeneity to ease notation):

$$f\left(P_{itk}^q|\mathbf{x}_{it}\right) = \sum_{r \in \Omega_{itk}} \Pr\left(R_{itk}^q = r|\mathbf{x}_{it}\right) \times \Pr\left(LB_{itk}^r \leq P_{itk}^{*q} < UB_{itk}^r|\mathbf{x}_{it}\right)$$

where $\Pr\left(LB_{itk}^r \leq P_{itk}^{*q} < UB_{itk}^r|\mathbf{x}_{it}\right)$ is given by

$$\Pr\left(LB_{itk}^r \leq P_{itk}^{*q} < UB_{itk}^r|\mathbf{x}_{it}\right) = \begin{cases} \Pr\left(LB_{itk}^r \leq P_{itk}^{*q}|\mathbf{x}_{it}\right) & \text{if } P_{itk}^q \geq P_{it,k-1}^q - 0.5r \\ \Pr\left(LB_{itk}^r \leq P_{itk}^{*q} < UB_{itk}^r|\mathbf{x}_{it}\right) & \text{if } 0.5r \leq P_{itk}^q < P_{it,k-1}^q - 0.5r \\ \Pr\left(P_{itk}^{*q} < UB_{itk}^r|\mathbf{x}_{it}\right) & \text{if } P_{itk}^q < 0.5r \end{cases}$$

All probabilities in the equation above are calculated from univariate normal distributions and are therefore easy to obtain. Note that whether a probability is censored or not depends on the degree of rounding and on the preceding probability.

5 Results

This section presents estimation results for the two models of subjective life expectancy explained above. The difference between the models is that the first one does not account for rounding, while the second model does. Descriptive statistics for all covariates used are given in Table B1 of Appendix B. In the main text we only report estimates for the equations that govern expectations. Estimates of the recall error and rounding processes can be found in Table C1 of Appendix C. The sample from which the estimates presented in the main text are obtained limits the data to complete and consistent responses for *both* sets of probabilities. Moreover, we only use the probabilities corresponding to those target ages for which *both* a PB and a DHS probability are available. Estimates based on all complete and consistent responses for either one of the datasets, regardless of whether the target age is included in both questionnaires, corroborate the findings from the main text and can be found in Appendix D.

5.1 Model without rounding of reported probabilities

Estimation results of the model without rounding are presented in the left panel of Table 3 (see section 4.1 for a detailed description of this model). The first two columns on the left present the effects of covariates on the baseline hazard as hazard ratios and the third column contains the differences between these hazard ratios across the two surveys. The estimated effects of most covariates are both qualitatively and quantitatively very similar for the PB and the DHS, with the exception of the cohort dummies. The baseline cohort 1942-1951 has a relatively low hazard of death according to the DHS: the hazard rates for the cohorts between 1952 and 1981 are between 15 and 30 percent higher than the baseline. However, according to the PB only the cohort 1952-1961 has a significantly higher hazard than the baseline and the difference is only 12 percent. These large differences between cohorts in the DHS and smaller and mostly insignificant differences in the PB lead us to reject the null hypotheses of equal cohort effects for all cohorts.

We do not find evidence to suggest that the two surveys generate different results for the other covariates. The dummy for the year 2012 is insignificant for both surveys. Women report a lower hazard of death compared to men, the hazard ratio is 93% according to the PB and 95% in the DHS. We find some disagreement between the PB and the DHS for the income dummy corresponding to a net household income of 1151-1800 euro per month. Based on the PB individuals in this group have a 18% higher hazard of death than the baseline of individuals in households that earn more than 2600 euro per month. However, in the DHS this difference does not exist. Such disagreement is not there for the other income groups, for which we cannot reject the null of equal coefficients. The education dummies show similar patterns for the PB and the DHS: respondents in the middle education category have a 14-16% lower hazard of death than their less educated peers. Though the PB shows a statistically significant difference of 9% for the high education category, this difference is only 2% and not significant for the DHS. However, the coefficient does not differ significantly between the surveys. As for self-reported health, respondents who rate their current health more positively report substantially lower hazards of death regardless of the set of probabilities used. The average hazard of respondents who rate their health as “not good” or “poor” is 86-94% higher than that of respondents who rate their health as “excellent”. None of the coefficients for the health variables differs significantly between the two surveys.

The overall picture that merges from the model that does not allow for rounding is that most correlations between covariates and expectations are similar in both datasets. However,

Table 3: Gompertz models of subjective survival

	Model 1 – No rounding			Model 2 – Rounding		
	PB[a]	DHS[a]	Diff. PB - DHS	PB[a]	DHS[a]	Diff. PB - DHS
Coh. 1932-41	1.128	0.975	0.153**	1.240***	0.888**	0.352***
	(0.0833)	(0.0646)	(0.0673)	(0.0672)	(0.0433)	(0.0589)
Coh. 1952-61	1.118**	1.276***	-0.158***	1.055	1.160***	-0.105***
	(0.0574)	(0.0607)	(0.0520)	(0.0354)	(0.0364)	(0.0389)
Coh. 1962-71	0.930*	1.147***	-0.217***	1.020	1.278***	-0.258***
	(0.0373)	(0.0559)	(0.0489)	(0.0360)	(0.0483)	(0.0452)
Coh. 1972-81	0.956	1.298***	-0.342***	1.120**	1.316***	-0.195***
	(0.0567)	(0.0831)	(0.0686)	(0.0501)	(0.0520)	(0.0558)
Coh. 1982-87	0.813	0.981	-0.168*	0.895	0.954	-0.0590
	(0.115)	(0.125)	(0.0931)	(0.104)	(0.0670)	(0.0737)
Wave 2012	1.009	0.993	0.0165	0.997	1.003	-0.00603
	(0.0236)	(0.0177)	(0.0272)	(0.0196)	(0.0151)	(0.0223)
Female	0.927**	0.948*	-0.0207	0.830***	0.904***	-0.0741***
	(0.0293)	(0.0290)	(0.0295)	(0.0225)	(0.0220)	(0.0244)
Net HH. Inc. $\leq$ €1150	0.980	0.928	0.0524	1.220***	1.094	0.126
	(0.0748)	(0.0752)	(0.0730)	(0.0891)	(0.0737)	(0.0830)
Net HH. Inc. €1151-1800	1.181***	0.994	0.188***	1.274***	1.046	0.228***
	(0.0522)	(0.0416)	(0.0521)	(0.0567)	(0.0372)	(0.0540)
Net HH. Inc. €1801-2600	0.933*	0.925**	0.00850	0.924***	0.938***	-0.0138
	(0.0332)	(0.0303)	(0.0336)	(0.0270)	(0.0229)	(0.0290)
Educ. middle	0.858***	0.838***	0.0202	1.025	0.958	0.0668*
	(0.0344)	(0.0334)	(0.0353)	(0.0363)	(0.0299)	(0.0360)
Educ. high	1.091***	1.024	0.0672	1.151***	1.057*	0.0936**
	(0.0369)	(0.0419)	(0.0413)	(0.0323)	(0.0327)	(0.0373)
Health: good	1.263***	1.346***	-0.0825	1.437***	1.303***	0.134***
	(0.0416)	(0.0554)	(0.0599)	(0.0407)	(0.0401)	(0.0512)
Health: fair	1.725***	1.710***	0.0156	2.153***	1.717***	0.436***
	(0.0755)	(0.0838)	(0.0923)	(0.0953)	(0.0691)	(0.0976)
Health: not good/poor	1.859***	1.938***	-0.0782	2.199***	2.001***	0.198
	(0.139)	(0.143)	(0.157)	(0.117)	(0.0977)	(0.133)
Constant	0.00650***	0.00526	0.00124	0.00531***	0.00436***	0.000950*
	(0.000335)	(0)	(0.000335)	(0.000310)	(0.000430)	(0.000499)
Chi2 test joint equality (16df)	86.90 ($p < 0.0001$)			154.25 ($p < 0.0001$)		
Chi2 test joint equality no cohorts (11df)	36.04 ($p = 0.0002$)			69.60 ($p < 0.0001$)		
Baseline hazard ($t/100$)	8.119***	8.084***	0.0342	8.104***	8.385***	-0.282**
	(0.0765)	(0.0775)	(0.0992)	(0.0696)	(0.123)	(0.140)
Variance ind. effects	0.771***	0.481***		0.635***	0.431***	
	(0.0400)	(0.0265)		(0.0248)	(0.0185)	
Corr. ind. effects	0.870***			0.787***		
	(0.0163)			(0.0155)		
Variance seq. effects	0.0818***	0.0610***		0.112***	0.0300***	
	(0.0153)	(0.0114)		(0.00776)	(0.00489)	
Corr. seq. effects	0.0324			0.239***		
	(0.0774)			(0.0743)		
Fraction var. ind. effects	0.904***	0.888***		0.851***	0.935***	
	(0.0175)	(0.0213)		(0.0107)	(0.0107)	
No. individuals		1,470			1,470	
No. probabilities		4,034			4,034	
Log-likelihood		-30,530.175			-16,048.925	

[a] Estimates reported as hazard ratios.
Standard errors in parentheses; $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

differences between birth cohorts are much larger in the DHS than in the PB. Moreover, we reject equality of coefficients for one income group. The Chi-squared tests for joint equality of coefficients across the PB and DHS reported in Table 3 reflect these observations: we reject the null of joint equality and much more strongly so if we take the cohort dummies into account.

The bottom of Table 3 reports other estimates. The baseline hazard is significant and positive for both datasets, which means that the hazard of death increases with age. Moreover, the estimated coefficients are very close, around 8.1 for both datasets, and the difference is not statistically significant. The estimated variances of the individual effects indicate that expectations are persistent at the level of the individual for both datasets: around 90% of the variance in expectations that cannot be explained by covariates is due to permanent unobserved heterogeneity. Furthermore, the individual effects are strongly positively correlated with a correlation coefficient of 0.87.

Table C1 in Appendix C presents estimates of the coefficients that capture heteroskedasticity of the recall error, capturing variation in the extent to which reported probabilities fit the Gompertz distribution. In addition to some differences between cohorts, the only factor that affects recall error similarly in both sets of probabilities is education. The middle and high education categories report probabilities that are significantly less noisy compared to respondents who have not finished vocational training.

Table D1 in Appendix D contains estimates of the exact same model, estimated on the larger sample of complete and consistent responses to either set of survival questions, using all available probabilities (also those target ages that are not included in one of the questionnaires). The same general picture emerges, but the differences between estimated cohort effects are smaller. Furthermore, we reject equality of coefficients for one additional income dummy (for an income between 1801 and 2600 euro per month).

5.2 Model with rounding of reported probabilities

Estimates for the model that accounts for rounding, described in section 4.2, are reported in the right panel of Table 3. As was the case without rounding, the model with rounding shows that the significant relationships between the hazard of death and covariates that emerge for the PB and the DHS have the same sign in almost all cases. The only exception is the oldest cohort, which has a 24% higher hazard than the baseline according to the PB but a 11% lower hazard based on the DHS. Moreover, the size of many correlations remains comparable between the surveys. However, incorporating rounding does not reduce the differences between the estimates from the two datasets and actually leads to more frequent rejections of equality. In addition to the dummy for household income between 1151 and 1800 euro per month, we also reject equality for the variables capturing gender and education and for two out of three indicators for health. Note that the finding that disparities between datasets are larger once we account for rounding can only occur in a model that point identifies beliefs. In the partial identification framework of section 3 rounding can only mitigate differences between imperfectly observed data.

The baseline hazard is similar across the PB and the DHS, and with a values of 8.1 and 8.4 duration dependence is similar to the values found in the model without rounding. For unobserved heterogeneity too the model with rounding corroborates the findings from that without rounding. Expectations are persistent at the level of the individual for both sets of probabilities. Question sequence effects are also significant, but much smaller in magnitude. Finally, Table 5 shows that the correlation between the individual effects for the PB and DHS questionnaires is 0.86, which is similar to that found in the baseline model.

The right panel in Table C1 contains the remaining estimates. The third and fourth column in Table C1 show the estimates for the heteroskedasticity of recall errors in the PB and DHS respectively. The variance of the errors is significantly lower among higher education groups, as was the case in the model without rounding. Compared to the left panel there is

Table 4: Model-implied average rounding probabilities

Multiples of...	Pension Barometer (%)	DNB Household Survey (%)
...100	1	2
...50	5	4
...25	11	–
...10	47	95
...5	33	–
...1	4	–

one additional column, which shows the estimated coefficients of the rounding equation for the PB. The estimates for the rounding equation in the DHS are not reported, because the thresholds for the rounding rule became arbitrarily large and standard errors could not be computed due to flatness of the simulated log-likelihood function. In other words: estimation strongly indicates that almost all probabilities in DHS are rounded to multiples of 10. The coefficients of the rounding equation for the PB, shown in the final column, also come with large standard errors. However, we do estimate the thresholds between different levels of rounding precisely. None of the 95 percent confidence intervals overlap, which indicates that we successfully identify the fractions of individuals that use different rounding rules. The sample average rounding probabilities are reported in Table 4, which shows that half of the reported probabilities are rounded to multiples of 10 and a third is rounded to multiples of 5. As suggested by the numerical issues associated with estimating the rounding equation for the DHS, 95 percent of probabilities reported in the DHS are rounded to multiples of 10.

5.3 Model fit

The results in the previous two subsections show that our conclusions regarding the reliability of subjective longevity are similar regardless of whether we account for rounding in our model of expectations. However, accounting for rounding does improve model fit. Figure 4 shows six histograms of reported probabilities in the data and of simulated probabilities from the models with and without rounding, pooling together all target ages. Even though the

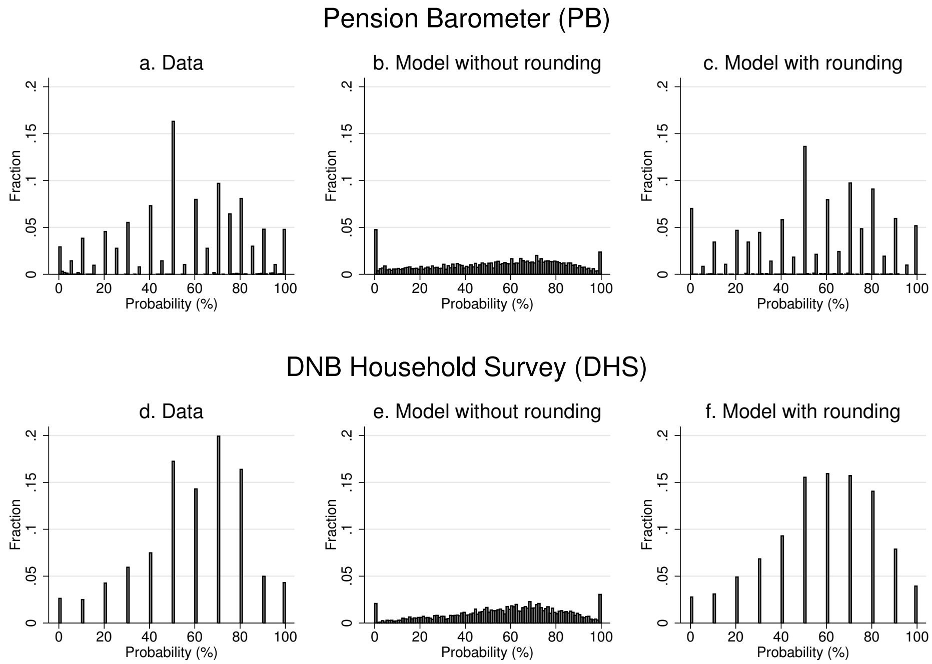

Figure 4: Histograms of data and simulated probabilities

PB allows respondents to report any probability between zero and one hundred, panel a. shows that resulting answers are bunched at multiples of 10. In fact, the lower part of the distribution, up to and including 50 percent, is similar to that of the DHS shown in panel d. The model without rounding cannot mimic such bunching, see panels b. and e., but the model that accounts for rounding does fit the data relatively closely (panels c. and f.). Hence, censoring at 0, 100 or the previous probability by itself does not produce the heaping at multiples of 10 that we observe in the data.

While the histograms in Figure 4 illustrate the importance of rounding, we may prefer to look at estimated densities in order to evaluate model fit. It is particularly difficult to compare the fit of the model with rounding and that without rounding, since the former is discrete while that latter is mostly continuous (except for the censoring). As a consequence, the model without rounding necessarily smooths the data more. Figure 5 displays estimated densities for the data and for simulated probabilities from both models. We find that the

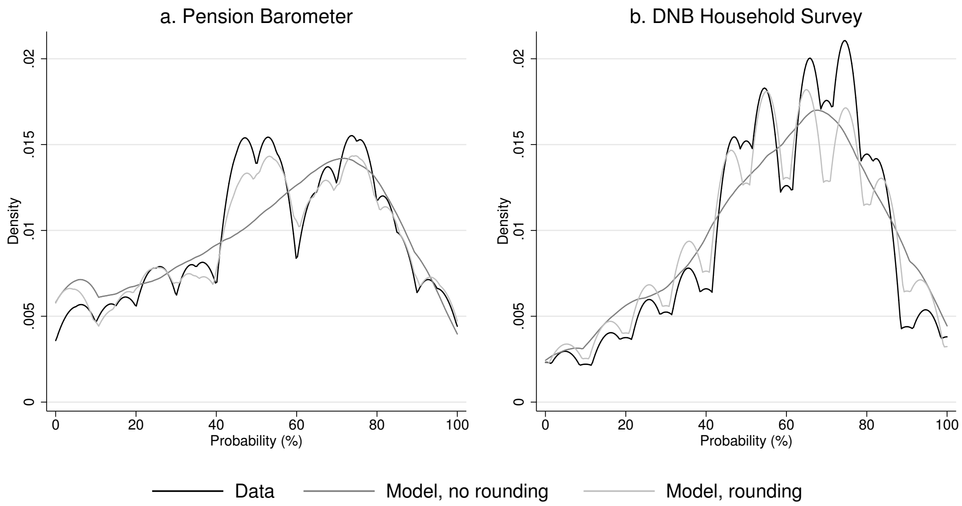

Figure 5: Kernel densities of data and simulated probabilities

density of the model without rounding fits the data much better than might be expected from the histograms: it provides a reasonable smoothed approximation of the bumpy density fitted on the data. This illustrates that even without rounding the model is fairly successful in distributing probability mass over the interval between 0 and 100, even if it does not place the mass at the limited set of probabilities that we observe in the data. The model that accounts for rounding does an even better job.

Comparing the log-likelihoods of the specifications in Table 3 with those of constant-only models reported in Appendix E, we find that covariates do not play an important role. The pseudo R-squared is around 0.006 for the model without rounding. Though many covariates correlate significantly with the hazard of death, most of the variation in expectations is explained by individual effects.

6 Subjective longevity in lifecycle models

In this section we evaluate whether subjective life expectancy is sufficiently reliable to be used as input for the estimation of empirical life-cycle models of saving and labor supply.

We use a standard life-cycle model to map reported probabilities in the PB and DHS into simulated wealth and labor supply profiles. This allows us to quantify the consequences of variation in subjective probabilities between surveys, their unreliability, in terms of economic outcomes. We proceed in three steps. First, we use the estimates reported in Table 3 to simulate the probability of dying in a given year conditional on current health. Two such sets of probabilities are computed, one for each survey. We then formulate a life-cycle model of saving and labor supply that links expectations to economic behavior. Finally, we compare the simulated wealth and labor supply profiles obtained using both sets of probabilities. The following subsections describe these steps in turn.

6.1 Simulating probabilities

The relevant input for a life-cycle model is the probability of dying at a certain age conditional on having survived to that age. We simulate probabilities for a male born between 1952 and 1961; with a net household income between 1800 and 2600 euros/month; and with a medium level of education. Since the life-cycle model uses a dichotomous measure of health, we alternatively fix health at "poor" and at "excellent". For covariates fixed at these values the relevant probabilities can be simulated from the estimates in Table 3 by integration over the distributions of individual and sequence effects. We use 10,000 draws of unobserved heterogeneity and simulate the relevant probabilities as the averages over those draws.[4]

Figure 6 plots the simulated probabilities of dying at different ages conditional on surviving to those ages. The left panel uses estimates from the model without rounding, while the right panel simulates probabilities from the model that takes rounding into account. Both panels show that the probability of death increases with age and that the increase is markedly stronger for people in poor health. Death at age 50 is extremely unlikely regardless of current

[4] The same draws are used for all ages.

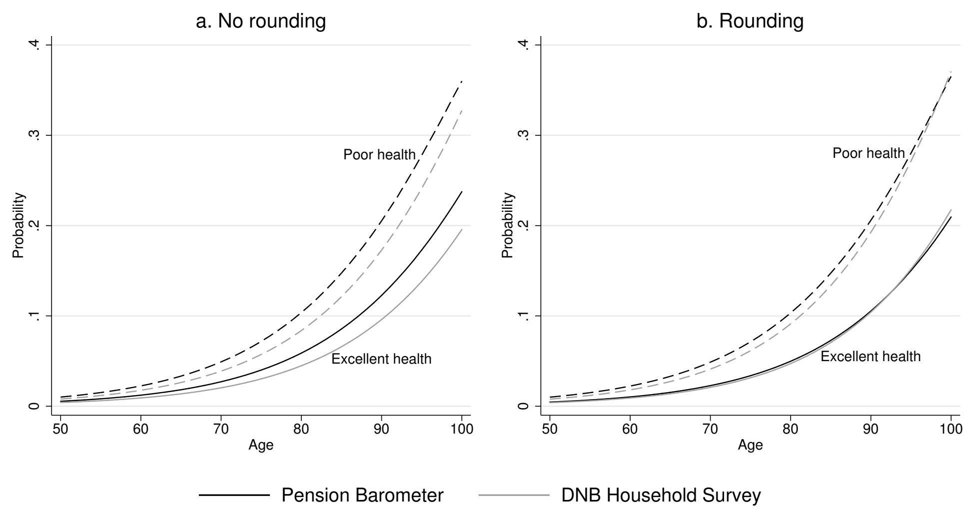

Figure 6: Simulated probabilities of dying at a given age conditional on survival up to that age

health, while above age 90 men in poor health face probabilities above 30% compared with 20% for those in excellent health.

Note that the differences between the PB and the DHS are larger for the model without rounding than for the model with rounding. This seems at odds with the finding that both parameters of the survival function are estimated to be more similar in the former model than in the latter (see Table 3 and accompanying discussion). The explanation is that the larger differences between the estimates for the γ and α-parameters for the model with rounding cancel out partly when combined into the probabilities shown in panel b. of Figure 6. The estimated baseline hazard is higher for the DHS than for the PB, but the γ-parameter through which covariates influence mortality is lower in the DHS (especially for "poor" health, see Table 3). In the model without rounding the difference between baseline hazards is much smaller, so differences between the γ-estimates translate directly into differences between probabilities (see panel a. of Figure 6).

6.2 A life-cycle model of saving and labor supply

We use a life-cycle model to translate the probabilities reported in Figure 6 into saving and labor supply decisions. The model is specified to approximate the institutions in place in the Netherlands in 2011/2012. We present the main features here and refer the reader to De Bresser et al. (2015) for a more detailed description.

The model is unitary: we assume a single decision maker per household. It spans the age-range 50 to 100 with a resolution of one year. Every year agents decide how much to save. Up to age 70 agents also choose their labor supply (0, 1500, 2000 or 2500 hours of work per year). There are three exit routes from the labor market: disability insurance (DI), unemployment insurance (UI) and occupational pensions. As long as they are in bad health agents may decide to claim disability insurance. Unemployment benefits can be claimed for a maximum of three years depending on one's work history. The level of unemployment and disability benefits is fixed at 70% of previous earnings. Moreover, both stop at age 65 when they are replaced by a flat-rate public pension. All 65-year olds receive the public pension regardless of their labor supply. If a worker's job includes an occupational pension plan the worker is obliged to participate. Occupational pensions can first be claimed at any age between 60 and 70 and benefits are a function of the number of years worked and average earnings. Occupational pension benefits are adjusted actuarially for the age at which they are first claimed. Agents cannot work while they receive DI, UI or occupational pensions. In addition to paid employment and the four types of transfers mentioned above, additional income is provided by an exogenous income stream generated by the partner. The state variables included in the model are wealth, wage, a binary health variable and eligibility indicators for DI, UI and occupational pensions.

There are three sources of uncertainty in the model: health, mortality and involuntary unemployment. The probability of being in "excellent" health next period is a function of current health and age. We estimate this health process using four waves of the SHARE-survey.[5] The mortality process consists of the probabilities reported in Figure 6. While of working age, agents face a risk of involuntary unemployment that we estimate from SHARE. The estimated constant risk of unemployment is 3% per year.

Agents with current age t derive utility from consumption c_t and leisure l_t according to the following utility function:

$$u\left(c_t, l_t, t\right) = n_t \frac{\left(\left(\frac{c_t}{n_t}\right)^{\kappa} l_t^{1-\kappa}\right)^{1-\sigma} - 1}{1-\sigma}$$

$$l_t = 4000 - h_t - \xi \mathbb{I}\left\{h_t > 0\right\} - \delta \mathbb{I}\left\{\text{bad health}\right\}$$
$$- \phi \mathbb{I}\left\{di_t > 0\right\} - \zeta \mathbb{I}\left\{ui_t > 0\right\}$$

n_t is an equivalence scale that reflects family size and that decreases with age; h_t is the number of hours worked; and all Greek letters denote parameters that are held constant in the simulations. The maximum amount of leisure is fixed at 4000 hrs/yr and bad health and claiming either UI or DI benefits carries stigma costs that are also measured in hrs/yr.

In addition to consumption and leisure agents value leaving behind a bequest according to the following bequest utility function:

$$b\left(w_t\right) = \exp\left(\theta_0 + \theta_1 t\right) \frac{\left(w_t + K\right)^{\kappa(1-\sigma)}}{1-\sigma}$$

where w_t is wealth at age t and K and the Greek symbols are parameters. The strength of this bequest motive depends on age, allowing it to vary with the size of the household as captured by the equivalence scale. This variation captures the notion that agents may care more about leaving wealth to their partner than to other individuals outside the household. Men who expect to outlive their partner care less about bequests than do men who expect to die younger.

[5] See http://www.share-project.org for more information on SHARE.

Table 5: Values for preference parameters used in simulations

Utility function		Leisure costs (hrs/yr)		Bequest utility	
σ – concavity	2.0	ξ – fixed cost of work	500	θ_0 – constant	-10.0
κ – consumption share	0.6	δ – cost of poor health	1000	θ_1 – HH size coefficient	7.9
β – discount factor	0.97	ϕ – stigma costs DI	3900	K – bequest concavity ($)	200,000
		ζ – stigma costs UI	3900		

The values of the parameters of both felicity functions are chosen to be in line with previous work where possible. Those parameters on which little prior evidence exists are set to produce reasonable wealth and labor supply profiles. All parameter values used in the simulations are shown in Table 5. Sensitivity analysis indicates that the findings described in the next subsection regarding the consequences of the reliability of subjective survival are robust to variation in the preference parameters.

6.3 Subjective survival and economic behavior

We use the life-cycle model described above to simulate wealth and labor supply for 5000 workers. Initial conditions are taken from SHARE (see De Bresser et al., 2015, for more information). The age profiles used to summarize simulations are those typically used to estimate preference parameters. For wealth, we compute quartiles at two-year age bins bins between ages 50 and 90. Labor supply is summarized by the average yearly hours worked by two-year bins for ages 50 to 70. In light of the dependence between mortality and health we calculate average labor supply conditional on current health.

Figure 7 presents our simulations. Results for the model that does not account for rounding are shown in the top row while results for the model that does take rounding into account are in the bottom row. The leftmost panels a. and d. contain the mortality processes discussed in section 6.1 and shown in Figure 6. The middle column, panels b. and e., displays wealth quartiles. We find that level of simulated wealth is sensitive to survival expectations. For the model without rounding, panel b., the maximum difference between the medians

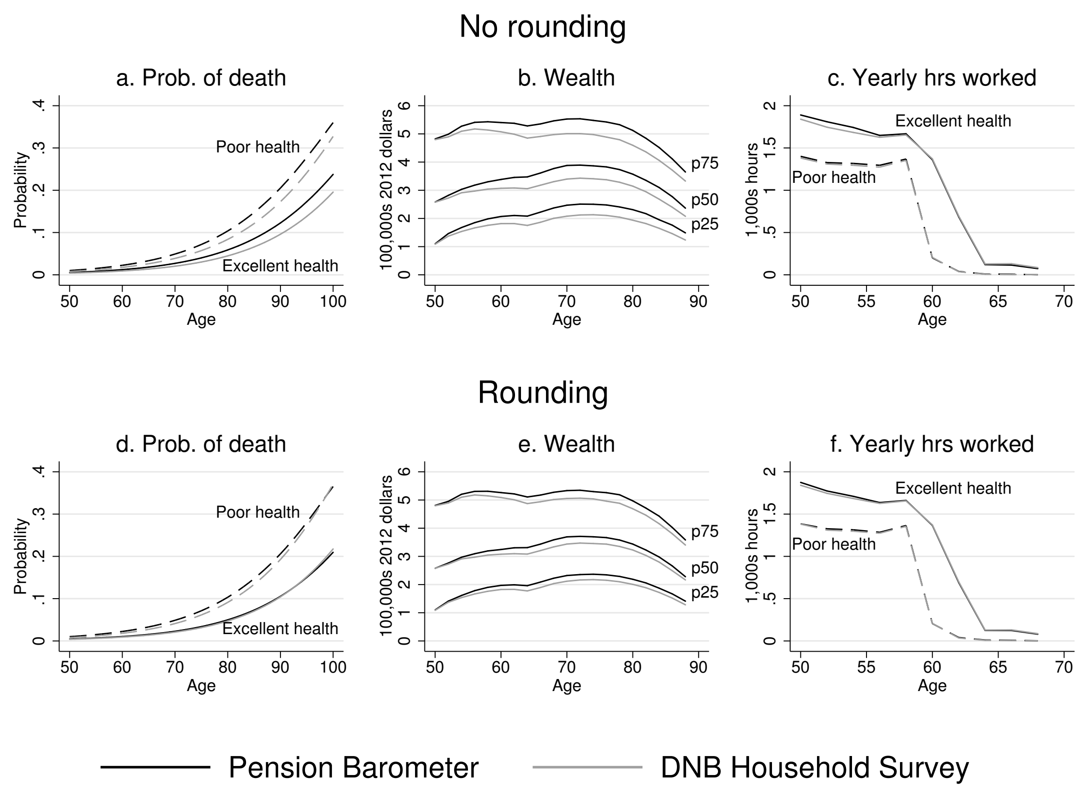

Figure 7: The impact of the reliability of subjective expectations on behavior in a life-cycle model

simulated based on the PB and the DHS is around 50,000 dollar or 13% of the PB profile. Simulated workers accumulate less wealth if we use the DHS mortality probabilities, despite the fact that they expect to live longer according to the DHS. Further simulations indicate that this is due to the assumption that the bequest motive is weaker at older ages. Men who expect to live longer have a weaker bequest motive and accumulate less wealth than those who expect to die younger.

The model that does account for rounding leads to smaller differences in mortality expectations between the PB and DHS (see panel d.). It therefore also leads to smaller differences between simulated wealth profiles: the maximum difference between the medians in panel e. is around 25,000 dollar or 7% of the PB profile. Hence, our simulations indicate that while the level of wealth is sensitive to the set of mortality probabilities used, the reliability of the expectations computed from the model with rounding is sufficient to lead to very similar wealth profiles.

Panels c. and f. show that labor supply is not sensitive to survival expectations. Regardless of whether we take rounding into account or not, the average hours worked per year is almost identical for the PB and the DHS. The reliability of subjective expectations appears adequate if the main focus of an analysis is on labor supply.

Varying the values for the preference parameters, we confirmed the robustness of the finding that wealth is more sensitive to expectations than labor supply. However, the extent of sensitivity of wealth does depend on the parameter values chosen. Stronger bequest motives lead to a reduction in sensitivity of wealth, reinforcing the notion that subjective expectations are sufficiently reliable to be used as inputs in structural models.

7 Conclusion

A growing body of research recognizes the potential of data that directly elicits expectations of survey respondents, so-called subjective expectations, especially in the context of inter-temporal models. However, many economists remain sceptical of the validity and informativeness of such data. This paper investigates the validity of reported expectations by evaluating the test-retest reliability of the type of expectations that has received most attention from researchers: survival expectations.

Using two surveys that were administered to the same respondents within the same month, we compare the answers of those respondents to items that ask for the likelihood of survival to various target ages. The questionnaires are the Pensioenbarometer (PB) and the DNB Household Survey (DHS), both of which were fielded to the CentERpanel, a household panel that is representative for the Dutch population. We take into account that the PB allows respondents to report any integer probability between 0 and 100 while the DHS limits responses to an 11-point scale between 0 and 10. We first analyze reliability at the level of the reported probability by checking whether reported probabilities are consistent with each other one-by-one. We check whether the rounded probabilities from both datasets are consistent with at least one underlying true probability under different degrees of rounding. We then analyze reported probabilities jointly in order to test whether the two surveys yield similar associations between expectations and background characteristics. This allows us to evaluate to what extent noise in the probabilities cancels out when those probabilities are combined in an aggregate model. We use the estimated models to construct survival curves and simulate labor supply and saving in a life-cycle model to quantify the economic implications of between-survey variation in elicited expectations.

We find the reliability of subjective survival expectations to be satisfactory overall. Test-retest correlations are in the 0.5-0.7 range, which is similar to the reliability of subjective well-being found by Krueger and Skade (2008). While around 20% of reported probabilities are equal in the PB and DHS, the fraction of consistent responses is much higher once we allow for rounding. Depending on the target age, 24-37% of reported probabilities are consistent if we assume that all PB probabilities are rounded to multiples of 1 and all DHS probabilities are rounded to multiples of 10. Common rounding as in Manski and Molinari (2010) raises the fraction of consistent probabilities to 32-46% and the most conservative degree of rounding for each reported probability increases it further to 61-77%.

Joint models of all reported probabilities show that both datasets yield quantitatively and qualitatively similar associations between socio-demographic covariates and the hazard of death. The largest differences between the estimates occur for cohort dummies. Other variables such as gender, income, education and self-assessed health enter the model in similar ways for both datasets, showing that reported expectations are reliable when probabilities are modelled jointly. We find that unobserved heterogeneity at the level of the individual is important and that this heterogeneity is strongly positively correlated across questionnaires.

We simulate saving and labor supply in a life-cycle model using survival curves constructed from the estimates of joint models of PB and DHS probabilities. The model with rounding yields more reliable survival curves than the model that does not take rounding into account. Saving is sensitive to survival expectations, to the extent that the variation in survival curves leads to a substantially different level of wealth if we construct curves from the estimates for the model without rounding (the difference is around 13%). However, the model that does take rounding into account halves the difference between wealth profiles to around 7%. Labor supply is less sensitive to survival expectations.

Taking all results together we conclude that the quality of subjective survival expectations is comparable to that of other types of subjective data that are frequently analyzed by economists, such as subjective well-being. Within-individual variation is both quantitatively less important and less reliable than variation between individuals, so applied researchers are advised not to focus exclusively on the former. When aggregated into survival curves, these data can be used to enrich inter-temporal models in which survival plays are role.

References

Bissonnette, L. and de Bresser, J. (2014). Eliciting subjective survival curves: Lessons from partial identification. Forthcoming in *Journal of Business and Economic Statistics*.

Bissonnette, L., Hurd, M. D., and Michaud, P.-C. (2011). *Individual survival curves combining subjective and actual mortality risks*. PhD thesis, Tilburg University. Appeared in PhD thesis titled “Essays on Subjective Expectations and Stated Preferences”.

De Bresser, J., Fonseca, R., and Michaud, P.-C. (2015). Retirement behavior in the U.S. and Europe. Mimeo, Tilburg University.

De Bresser, J. and Van Soest, A. (2013). Survey response in probabilistic questions and its impact on inference. *Journal of Economic Behavior and Organization*, 96(December 2013):65–84.

De Nardi, M., French, E., and Jones, J. B. (2010). Why do the elderly save? the role of medical expenditures. *Journal of Political Economy*, 118(1):39–75.

French, E. (2005). The effects of health, wealth and wages on labour supply and retirement behaviour. *Review of Economic Studies*, 72(2):395–427.

French, E. and Jones, J. B. (2011). The effects of health insurance and self-insurance on retirement behavior. *Econometrica*, 79(3):693–732.

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

Hurd, M. and McGarry, K. (1995). Evaluation of the subjective probabilities of survival in the Health and Retirement Study. *The Journal of Human Resources*, 30(Special Issue on the Health and Retirement Study: Data Quality and Early Results):S268–S292.

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

Krueger, A. B. and Skade, D. A. (2008). The reliability of subjective well-being measures. *Journal of Public Economics*, 92(8-9):1833–1845.

Kutlu, V. and Kalwij, A. S. (2012). Individuals' survival expectations and actual mortality. Utrecht University working paper.

Manski, C. F. (2004). Measuring expectations. *Econometrica*, 75(5):1329–1376.

Manski, C. F. and Molinari, F. (2010). Rounding probabilistic expectations in surveys. *Journal of Business and Economic Statistics*, 28(2):219–231.

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

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

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

Teppa, F., Thorp, S., and Bateman, H. (2015). Family, friends and framing: A cross-country study of subjective survival expectations. De Nederlandsche Bank Working Paper.

Appendix A Incidence of rounding

Table A1a: Common rounding

	PB (%)	DHS (%)
All 0 or 100	1	3
All 0, 50 or 100	2	3
All multiples of 10	23	95
All multiples of 5	58	
Some in $[1, 4]$ or $[96, 100]$	11	
Other	5	
Total	100%	100%

$N = 2,187$ individual-year observations

Table A1b: General rounding

Multiples of...	PB (%)	DHS (%)
...100	8	7
...50	16	17
...25	9	
...10	52	76
...5	12	
...1	3	
Total	100%	100%

$N = 4,062$ probabilities

Appendix B Descriptive statistics of covariates

Table B1: Descriptive statistics

	Probs. from PB *and* DHS		Probs. from PB *or* DHS	
	Mean	Std. dev.	Mean	Std. dev.
Coh. 1922-1931	–	–	0.04	0.19
Coh. 1932-1941	0.13	0.34	0.13	0.33
Coh. 1942-1951	0.28	0.45	0.27	0.45
Coh. 1952-1961	0.24	0.43	0.23	0.42
Coh. 1962-1971	0.21	0.41	0.18	0.39
Coh. 1972-1981	0.11	0.32	0.13	0.34
Coh. 1982-1987	0.02	0.14	0.02	0.14
Wave 2012	0.48	0.50	0.51	0.50
Female	0.43	0.50	0.44	0.50
Net HH. inc. $\leq$ €1150	0.06	0.24	0.08	0.27
Net HH. inc. €1151-1800	0.16	0.36	0.15	0.36
Net HH. inc. €1801-2600	0.28	0.45	0.24	0.43
Net HH. inc. $\geq$ €2601	0.51	0.50	0.53	0.50
Educ. low	0.29	0.45	0.30	0.46
Educ. middle	0.30	0.46	0.28	0.45
Educ. high	0.42	0.49	0.41	0.49
Health: excellent	0.14	0.34	0.14	0.34
Health: good	0.63	0.48	0.62	0.49
Health: fair	0.17	0.37	0.18	0.38
Health: not good/poor	0.07	0.26	0.07	0.25
N (individuals)	1,470		2,323	
N (individual-years)	2,073		3,787	

Appendix C Estimates of recall error and rounding equations

Table C1: Recall error and rounding estimates of Gompertz models of subjective survival

	Model 1 – No rounding		Model 2 – Rounding		
	Error PB	Error DHS	Error PB	Error DHS	Rounding PB
Coh. 1932-41	0.216***	0.198***	0.151*	0.372***	-0.0583
	(0.0677)	(0.0619)	(0.0771)	(0.0610)	(0.128)
Coh. 1952-61	0.0863	-0.0411	0.0726	-0.0269	-0.000522
	(0.0543)	(0.0482)	(0.0555)	(0.0496)	(0.0949)
Coh. 1962-71	0.0211	-0.142***	0.0434	-0.0521	-0.0438
	(0.0533)	(0.0525)	(0.0604)	(0.0531)	(0.103)
Coh. 1972-81	0.201***	0.102*	0.123*	-0.0788	-0.0763
	(0.0661)	(0.0613)	(0.0732)	(0.0641)	(0.120)
Coh. 1982-87	0.155	-0.194	-0.296	-0.605***	0.152
	(0.130)	(0.129)	(0.189)	(0.196)	(0.234)
Wave 2012	-0.0281	0.000787	-0.251***	0.0517	0.0325
	(0.0468)	(0.0395)	(0.0566)	(0.0442)	(0.0613)
Female	-0.0765**	0.0226	0.0413	0.0916**	0.0326
	(0.0373)	(0.0344)	(0.0428)	(0.0371)	(0.0710)
Net HH. Inc. $\leq$ €1150	-0.0227	0.292***	0.146	0.208**	-0.156
	(0.0791)	(0.0847)	(0.0979)	(0.0829)	(0.153)
Net HH. Inc. €1151-1800	0.0289	0.00558	0.251***	-0.0297	-0.185*
	(0.0526)	(0.0515)	(0.0604)	(0.0557)	(0.104)
Net HH. Inc. €1801-2600	0.0596	-0.0724*	0.237***	-0.140***	-0.0218
	(0.0444)	(0.0411)	(0.0522)	(0.0464)	(0.0814)
Educ. middle	-0.123**	-0.180***	-0.146**	-0.272***	0.0348
	(0.0541)	(0.0485)	(0.0574)	(0.0495)	(0.0966)
Educ. high	-0.247***	-0.152***	-0.259***	-0.221***	-0.0714
	(0.0503)	(0.0451)	(0.0525)	(0.0452)	(0.0930)
Health: good	0.0830	-0.0962	0.0838	-0.0350	0.00927
	(0.0853)	(0.0643)	(0.0682)	(0.0571)	(0.115)
Health: fair	0.174*	0.00757	0.209***	0.167**	-0.311**
	(0.0990)	(0.0748)	(0.0802)	(0.0678)	(0.138)
Health: not good/poor	0.0333	0.314***	0.201*	0.101	-0.148
	(0.121)	(0.100)	(0.104)	(0.0901)	(0.176)
Constant	2.363***	2.586***	2.105***	2.337***	
	(0.118)	(0.0795)	(0.0894)	(0.0718)	
μ_1					-2.491***
					(0.169)
μ_2					-0.584***
					(0.167)
μ_3					1.154***
					(0.182)
μ_4					1.973***
					(0.199)
μ_5					3.088***
					(0.240)
Variance ind. effects					0.693***
					(0.0995)
Variance seq. effects					0.0284*
					(0.0157)
No. individuals	1,470		1,470		
No. probabilities	4,034		4,034		
Log-likelihood	-30,530.175		-16,048.925		

Standard errors in parentheses; ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Appendix D Estimates based on all valid probabilities

Table D1: Gompertz model of subjective survival – estimates based on all valid probabilities

	Model 1 – No rounding			Model 2 – Rounding		
	PB[a]	DHS[a]	Diff. PB - DHS	PB[a]	DHS[a]	Diff. PB - DHS
Coh. 1922-31	1.145*	1.171*	-0.0265	0.985	0.971	0.0147
	(0.0932)	(0.0999)	(0.104)	(0.0709)	(0.0768)	(0.0817)
Coh. 1932-41	1.053	1.136**	-0.0825	0.962	1.057	-0.0948**
	(0.0612)	(0.0634)	(0.0518)	(0.0371)	(0.0420)	(0.0447)
Coh. 1952-61	1.036	1.072**	-0.0365	0.926***	1.103***	-0.177***
	(0.0326)	(0.0360)	(0.0378)	(0.0219)	(0.0275)	(0.0294)
Coh. 1962-71	0.928**	0.997	-0.0696**	0.925**	1.123***	-0.198***
	(0.0322)	(0.0363)	(0.0340)	(0.0305)	(0.0336)	(0.0318)
Coh. 1972-81	0.777***	0.869**	-0.0920**	1.067**	1.179***	-0.112***
	(0.0561)	(0.0531)	(0.0371)	(0.0344)	(0.0352)	(0.0385)
Coh. 1982-87	1.114	0.951	0.163	1.216***	1.060	0.156**
	(0.155)	(0.108)	(0.104)	(0.0810)	(0.0491)	(0.0730)
Wave 2012	1.012	1.008	0.00365	1.022	1.035***	-0.0128
	(0.0186)	(0.0161)	(0.0212)	(0.0143)	(0.0122)	(0.0176)
Female	1.013	1.028	-0.0152	0.852***	0.890***	-0.0375*
	(0.0267)	(0.0253)	(0.0273)	(0.0166)	(0.0178)	(0.0198)
Net HH. Inc. $\leq$ €1150	1.108	1.031	0.0771	1.009	1.062	-0.0532
	(0.0766)	(0.0507)	(0.0737)	(0.0580)	(0.0414)	(0.0607)
Net HH. Inc. €1151-1800	1.046	0.940*	0.107***	1.024	0.990	0.0340
	(0.0410)	(0.0333)	(0.0384)	(0.0370)	(0.0283)	(0.0379)
Net HH. Inc. €1801-2600	1.039	0.966	0.0736**	1.017	1.025	-0.00787
	(0.0273)	(0.0236)	(0.0295)	(0.0223)	(0.0216)	(0.0265)
Educ. middle	0.852***	0.824***	0.0288	0.908***	0.904***	0.00386
	(0.0356)	(0.0321)	(0.0283)	(0.0239)	(0.0220)	(0.0266)
Educ. high	0.995	1.017	-0.0219	0.883***	0.912***	-0.0288
	(0.0312)	(0.0318)	(0.0301)	(0.0160)	(0.0204)	(0.0226)
Health: good	1.212***	1.210***	0.00245	1.361***	1.230***	0.131***
	(0.0387)	(0.0397)	(0.0357)	(0.0236)	(0.0266)	(0.0322)
Health: fair	1.719***	1.618***	0.101	1.975***	1.613***	0.362***
	(0.0680)	(0.0657)	(0.0655)	(0.0606)	(0.0485)	(0.0673)
Health: not good/poor	2.044***	1.858***	0.186	2.183***	1.817***	0.366***
	(0.116)	(0.105)	(0.114)	(0.0885)	(0.0833)	(0.103)
Constant	0.00310***	0.0222***	-0.0191***	0.00307***	0.0188***	-0.0157***
	(0.000103)	(0.00157)	(0.00160)	(8.70e-05)	(0.000815)	(0.000820)
Chi2 test joint equality (17df)	203.25 ($p < 0.0001$)			577.95 ($p < 0.0001$)		
Chi2 test joint equality no cohorts (11df)	161.24 ($p < 0.0001$)			411.33 ($p < 0.0001$)		
Baseline hazard ($t/100$)	9.091***	6.211***	2.880***	9.123***	6.480***	2.643***
	(0.0680)	(0.0690)	(0.114)	(0.0344)	(0.0667)	(0.0742)
Variance ind. effects	0.809***	0.505***		0.850***	0.437***	
	(0.0413)	(0.0312)		(0.0256)	(0.0159)	
Corr. ind. effects	0.834***			0.781***		
	(0.0393)			(0.0115)		
Variance seq. effects	0.106***	0.0350***		0.104***	0.0234***	
	(0.00647)	(0.0131)		(0.00457)	(0.00346)	
Corr. seq. effects	0.442***			0.604***		
	(0.123)			(0.0644)		
Fraction var. ind. effects	0.884***	0.935***		0.891***	0.949***	
	(0.00696)	(0.0214)		(0.00521)	(0.00759)	
No. individuals		2,323			2,323	
No. probabilities		16,540			16,540	
Log-likelihood		-74,126.826			-40,588.262	

[a] Estimates reported as hazard ratios.
Standard errors in parentheses; $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table D2: Gompertz model of subjective survival – estimates based on all valid probabilities

	Model 1 – No rounding		Model 2 – Rounding		
	Error PB	Error DHS	Error PB	Error DHS	Rounding PB
Coh. 1922-31	-0.105	0.348***	-0.166*	0.368***	0.108
	(0.0740)	(0.0825)	(0.0853)	(0.0853)	(0.151)
Coh. 1932-41	-0.0216	0.251***	-0.0265	0.311***	0.101
	(0.0291)	(0.0519)	(0.0356)	(0.0484)	(0.0707)
Coh. 1952-61	0.0594***	-0.0159	0.0212	-0.0638	0.0162
	(0.0225)	(0.0412)	(0.0268)	(0.0390)	(0.0550)
Coh. 1962-71	-0.0260	-0.0391	-0.0280	-0.0604	-0.0232
	(0.0240)	(0.0414)	(0.0297)	(0.0402)	(0.0590)
Coh. 1972-81	0.0931***	0.0569	0.0589	-0.161***	0.0407
	(0.0294)	(0.0441)	(0.0360)	(0.0444)	(0.0690)
Coh. 1982-87	0.0260	-0.139	-0.266***	-0.435***	0.0630
	(0.0593)	(0.0903)	(0.0758)	(0.0926)	(0.126)
Wave 2012	-0.0221	0.0484*	-0.0514**	9.60e-05	-0.0330
	(0.0184)	(0.0257)	(0.0232)	(0.0293)	(0.0334)
Female	0.0512***	-0.00434	0.0316	0.0325	0.0503
	(0.0169)	(0.0236)	(0.0207)	(0.0256)	(0.0404)
Net HH. Inc. $\leq$ €1150	0.148***	0.204***	0.179***	0.140***	0.0804
	(0.0358)	(0.0488)	(0.0460)	(0.0538)	(0.0877)
Net HH. Inc. €1151-1800	0.0896***	0.105***	0.168***	0.0291	-0.117**
	(0.0251)	(0.0361)	(0.0303)	(0.0391)	(0.0586)
Net HH. Inc. €1801-2600	0.0350*	0.0124	0.0854***	-0.0569*	-0.0733
	(0.0198)	(0.0288)	(0.0248)	(0.0319)	(0.0473)
Educ. middle	-0.0410*	-0.0833**	-0.0990***	-0.117***	0.0967*
	(0.0223)	(0.0352)	(0.0279)	(0.0340)	(0.0547)
Educ. high	-0.192***	-0.154***	-0.224***	-0.126***	0.0119
	(0.0208)	(0.0338)	(0.0263)	(0.0317)	(0.0532)
Health: good	0.00435	0.00544	0.0157	0.0328	0.0277
	(0.0261)	(0.0488)	(0.0313)	(0.0387)	(0.0582)
Health: fair	0.0297	0.148***	0.0930**	0.272***	-0.128*
	(0.0316)	(0.0564)	(0.0384)	(0.0469)	(0.0721)
Health: not good/poor	0.0230	0.250***	0.0413	0.335***	-0.0670
	(0.0413)	(0.0734)	(0.0522)	(0.0702)	(0.0974)
Constant	2.550***	2.479***	2.404***	2.311***	
	(0.0331)	(0.0767)	(0.0408)	(0.0523)	
μ_1					-1.985***
					(0.0854)
μ_2					-0.374***
					(0.0855)
μ_3					1.271***
					(0.0914)
μ_4					1.981***
					(0.101)
μ_5					3.124***
					(0.134)
Variance ind. effects					0.440***
					(0.0476)
Variance seq. effects					0.0253***
					(0.00813)
No. individuals	2,323		2,323		
No. probabilities	16,540		16,540		
Log-likelihood	-74,126.826		-40,588.262		

Standard errors in parentheses; ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Appendix E Estimates from constant-only models

Table E1: Gompertz model of subjective survival – estimates of constant-only models

a. Common probabilities that are in complete and logically consistent sequences for both surveys

	Model 1 – no rounding			Model 2 – rounding		
	PB	DHS	Diff. PB - DHS	PB	DHS	Diff. PB - DHS
Gamma[a]	0.00378***	0.0124***	-0.00858***			
	(0.000228)	(0.00139)	(0.00138)			
Alpha	9.225***	7.367***	1.858***			
	(0.0818)	(0.156)	(0.169)			
Log SD errors	2.356***	2.451***				
	(0.0169)	(0.0166)				
Variance ind. effects	0.773***	0.560***				
	(0.0331)	(0.0241)				
Corr. ind. effects	0.852***					
	(0.0129)					
Variance seq. effects	0.0971***	0.0340***				
	(0.00718)	(0.00713)				
Corr. seq. effects	0.0550					
	(0.0781)					
Fraction var. ind. effects	0.888***	0.943***				
	(0.00840)	(0.0117)				
No. individuals		1,470			1,470	
No. probabilities		4,034			4,034	
Log-likelihood		-30,711.080				

b. All probabilities that are in complete and logically consistent sequences in either survey

	Model 1 – no rounding			Model 2 – rounding		
	PB	DHS	Diff. PB - DHS	PB	DHS	Diff. PB - DHS
Gamma[a]	0.00414***	0.0218***	-0.0177***			
	(0.000139)	(0.00156)	(0.00156)			
Alpha	9.120***	6.570***	2.550***			
	(0.0441)	(0.0995)	(0.107)			
Log SD errors	2.541***	2.522***				
	(0.00806)	(0.0111)				
Variance ind. effects	0.994***	0.602***				
	(0.0304)	(0.0348)				
Corr. ind. effects	0.829***					
	(0.0197)					
Variance seq. effects	0.0929***	0.0296***				
	(0.00493)	(0.00884)				
Corr. seq. effects	0.323***					
	(0.0925)					
Fraction var. ind. effects	0.915***	0.953***				
	(0.00491)	(0.0136)				
No. individuals		2,323			2,323	
No. probabilities		16,540			16,540	
Log-likelihood		-74,530.708				

[a] Estimates reported as hazard ratios.
Standard errors in parentheses; ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Appendix F Estimates of models of *remaining* lifetime

In the main text we model *total* subjective lifetimes, from birth to death, and condition on the current age of the respondent. Alternatively, we may also specify Gompertz distributions over the *remaining* lifespan from the current age of the respondent onwards. The latter approach is similar to that of fitting individual survival functions to the probabilities reported by survey respondents. This approach has been followed by several previous researchers, such as Perozek (2008). However, they estimate both parameters, α and γ, for each individual, while we estimate a proportional hazard model with α fixed and proportional effects of covariates on the baseline hazard. In our proportional hazard framework we prefer to model total rather than remaining lifetime, because the latter implies implausible features of the baseline hazard. In particular, it implies that the ratio of the hazards of surviving another five years to the hazard of surviving ten more years is the same for respondents with the same levels of covariates. This is not plausible given that we group birth cohorts in intervals of 10 years. Nonetheless, we report the estimates of an analogous analysis to that in the main text conducted on *remaining* rather than *total* lifetime to allow the reader to assess the robustness of our findings. In the model of remaining lifetime, true survival probabilities on a scale from 0 to 100 are given by:

$$S^q_{itk} = \exp\left(-\frac{\gamma^q_{it}}{\alpha^q}\left(\exp\left(\alpha^q\left(ta_k - a_{it}\right)\right) - 1\right)\right) \times 100$$

where q indexes questionnaires ($q \in \{PB, DHS\}$); $\gamma^q_{it} = \exp\left(\mathbf{x}'_{it}\boldsymbol{\beta}^q_1 + \xi^q_i + \eta^q_{it}\right)$ depends on the demographics of respondent i in survey-year t; α^q determines the shape of the baseline hazard; ta_k is a target age in the questionnaire and a_{it} is the age of i in year t. All other parts of the model are the same as for the specification for total lifetime explained in the text.

Table F1: Gompertz model of *remaining* subjective survival without rounding – model estimated on probabilities that were reported in both surveys

	PB[a]	DHS[a]	Diff. PB - DHS	Error PB	Error DHS
Coh. 1932-41	2.471***	2.138***	0.333**	0.174*	0.207***
	(0.155)	(0.134)	(0.137)	(0.0959)	(0.0719)
Coh. 1952-61	0.512***	0.572***	-0.0599**	0.131**	-0.0726
	(0.0270)	(0.0297)	(0.0243)	(0.0658)	(0.0517)
Coh. 1962-71	0.203***	0.231***	-0.0283**	0.0272	-0.128**
	(0.0106)	(0.0122)	(0.0112)	(0.0768)	(0.0588)
Coh. 1972-81	0.119***	0.135***	-0.0163*	0.250***	0.0556
	(0.00905)	(0.00939)	(0.00934)	(0.0791)	(0.0659)
Coh. 1982-87	0.0473***	0.0513***	-0.00399	0.157	-0.217*
	(0.0105)	(0.00868)	(0.00702)	(0.150)	(0.125)
Wave 2012	1.074***	1.061***	0.0133	-0.0402	0.00733
	(0.0264)	(0.0193)	(0.0288)	(0.0574)	(0.0414)
Female	0.911***	0.980	-0.0694**	0.00392	-0.0155
	(0.0329)	(0.0320)	(0.0293)	(0.0366)	(0.0356)
Net HH. inc. $\leq$ €1150	0.951	0.863*	0.0873	-0.178**	0.327***
	(0.0714)	(0.0689)	(0.0636)	(0.0798)	(0.0859)
Net HH. inc. €1151-1800	1.115***	0.981	0.134***	-0.117*	0.0184
	(0.0466)	(0.0405)	(0.0427)	(0.0674)	(0.0509)
Net HH. inc. €1801-2600	0.987	0.954	0.0335	-0.0108	-0.0557
	(0.0383)	(0.0379)	(0.0339)	(0.0655)	(0.0433)
Educ. middle	0.883***	0.855***	0.0276	-0.0817	-0.212***
	(0.0384)	(0.0367)	(0.0349)	(0.0735)	(0.0501)
Educ. high	0.865***	0.889***	-0.0242	-0.243***	-0.150***
	(0.0326)	(0.0381)	(0.0342)	(0.0832)	(0.0493)
Health: good	1.229***	1.241***	-0.0117	0.0873	-0.0823
	(0.0411)	(0.0627)	(0.0534)	(0.242)	(0.109)
Health: fair	1.593***	1.530***	0.0631	0.227	-0.0188
	(0.0709)	(0.0831)	(0.0826)	(0.245)	(0.113)
Health: not good/poor	1.653***	1.731***	-0.0785	0.0696	0.222
	(0.111)	(0.144)	(0.120)	(0.245)	(0.136)
Constant	0.0139***	0.0105***	0.00335***	2.337***	2.586***
	(0.000507)	(0.000803)	(0.000807)	(0.353)	(0.136)
Chi2 test joint equality (16df)	131.68*** ($p < 0.0001$)				
Chi2 test joint equality no cohorts (11df)	51.45*** ($p < 0.0001$)				
Baseline hazard	0.0746***	0.0793***			
	(0.00286)	(0.00238)			
Variance ind. effects	0.888***	0.553***			
	(0.0603)	(0.0367)			
Corr. ind. effects	0.877***				
	(0.0151)				
Variance seq. effects	0.0831***	0.0712***			
	(0.00815)	(0.0120)			
Corr. seq. effects	0.192***				
	(0.0631)				
No. individuals			1,470		
No. probabilities			4,034		
Log-likelihood			-30,577.676		

[a] Estimates reported as hazard ratios.
Standard errors in parentheses; ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Table F2: Gompertz model of *remaining* subjective survival without rounding – model estimated on all valid probabilities

	PB[a]	DHS[a]	Diff. PB - DHS	Error PB	Error DHS
Coh. 1922-31	5.409***	3.474***	1.935***	-0.0853	0.354***
	(0.647)	(0.354)	(0.508)	(0.0745)	(0.0809)
Coh. 1932-41	3.010***	2.248***	0.762***	-0.0247	0.252***
	(0.198)	(0.119)	(0.152)	(0.0293)	(0.0471)
Coh. 1952-61	0.417***	0.584***	-0.167***	0.0671***	-0.0323
	(0.0167)	(0.0211)	(0.0190)	(0.0222)	(0.0380)
Coh. 1962-71	0.137***	0.269***	-0.132***	-0.0278	-0.0350
	(0.00455)	(0.0106)	(0.00962)	(0.0237)	(0.0382)
Coh. 1972-81	0.0548***	0.138***	-0.0829***	0.104***	0.0533
	(0.00367)	(0.00719)	(0.00607)	(0.0293)	(0.0407)
Coh. 1982-87	0.0279***	0.0711***	-0.0432***	-0.00413	-0.0918
	(0.00377)	(0.00747)	(0.00583)	(0.0574)	(0.0797)
Wave 2012	1.068***	1.061***	0.00743	-0.0192	0.0378
	(0.0188)	(0.0145)	(0.0224)	(0.0185)	(0.0249)
Female	0.850***	0.908***	-0.0588**	0.0500***	0.00374
	(0.0252)	(0.0244)	(0.0244)	(0.0168)	(0.0229)
Net HH. inc. $\leq$ €1150	0.960	0.961	-0.00123	0.148***	0.214***
	(0.0496)	(0.0439)	(0.0587)	(0.0353)	(0.0481)
Net HH. inc. €1151-1800	0.968	0.898***	0.0694*	0.0782***	0.107***
	(0.0349)	(0.0313)	(0.0371)	(0.0248)	(0.0356)
Net HH. inc. €1801-2600	1.024	0.944**	0.0800***	0.0313	-0.00122
	(0.0264)	(0.0236)	(0.0287)	(0.0198)	(0.0285)
Educ. middle	0.899***	0.872***	0.0271	-0.0619***	-0.0775**
	(0.0309)	(0.0296)	(0.0298)	(0.0224)	(0.0330)
Educ. high	0.989	1.000	-0.0109	-0.215***	-0.147***
	(0.0250)	(0.0293)	(0.0312)	(0.0212)	(0.0307)
Health: good	1.241***	1.206***	0.0346	0.00310	-0.000129
	(0.0274)	(0.0323)	(0.0374)	(0.0261)	(0.0402)
Health: fair	1.648***	1.544***	0.104	0.0352	0.136***
	(0.0596)	(0.0555)	(0.0670)	(0.0319)	(0.0482)
Health: not good/poor	1.820***	1.702***	0.118	0.0258	0.231***
	(0.0966)	(0.0889)	(0.104)	(0.0415)	(0.0657)
Constant	0.0123***	0.0132***	-0.000932	2.564***	2.488***
	(0.000361)	(0.000518)	(0.000591)	(0.0328)	(0.0546)
Chi2 equality (17df)	382.37*** ($p < 0.0001$)				
Chi2 equality no cohorts (11df)	20.03** ($p = 0.0449$)				
Baseline hazard	0.0901***	0.0629***			
	(0.000832)	(0.00139)			
Variance ind. effects	0.940***	0.572***			
	(0.0302)	(0.0218)			
Corr. ind. effects	0.844***				
	(0.0110)				
Variance seq. effects	0.0971***	0.0277***			
	(0.00524)	(0.00767)			
Corr. seq. effects	0.321***				
	(0.0902)				
No. individuals			2,323		
No. probabilities			16,540		
Log-likelihood			-74,241.347		

[a] Estimates reported as hazard ratios.
Standard errors in parentheses; ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Table F3: Gompertz model of *remaining* subjective survival with rounding – model estimated on probabilities that were reported in both surveys

	PB[a]	DHS[a]	Diff. PB - DHS	Error PB	Error DHS	Rounding PB
Coh. 1932-41	2.225***	1.864***	0.361***	0.251***	0.374***	-0.0987
	(0.166)	(0.109)	(0.122)	(0.0771)	(0.0621)	(0.130)
Coh. 1952-61	0.453***	0.535***	-0.0822***	0.0852	-0.0755	0.109
	(0.0155)	(0.0149)	(0.0169)	(0.0542)	(0.0508)	(0.0956)
Coh. 1962-71	0.176***	0.190***	-0.0137*	0.0848	-0.109**	-0.0979
	(0.00681)	(0.00628)	(0.00710)	(0.0592)	(0.0541)	(0.101)
Coh. 1972-81	0.0986***	0.108***	-0.00918	0.0463	-0.0498	0.0206
	(0.00475)	(0.00497)	(0.00561)	(0.0794)	(0.0642)	(0.117)
Coh. 1982-87	0.0554***	0.0435***	0.0120***	-0.458***	-0.603***	0.261
	(0.00257)	(0.00302)	(0.00342)	(0.169)	(0.172)	(0.204)
Wave 2012	1.071***	1.094***	-0.0223	-0.124**	-0.00967	-0.00374
	(0.0193)	(0.0164)	(0.0237)	(0.0525)	(0.0459)	(0.0614)
Female	0.745***	0.830***	-0.0853***	0.0254	0.0625*	0.102
	(0.0203)	(0.0196)	(0.0213)	(0.0430)	(0.0378)	(0.0720)
Net HH. Inc. $\leq$ €1150	1.021	1.005	0.0162	-0.0233	0.354***	-0.204
	(0.0501)	(0.0557)	(0.0641)	(0.0895)	(0.0824)	(0.155)
Net HH. Inc. €1151-1800	1.157***	1.000	0.157***	0.214***	0.0658	-0.140
	(0.0505)	(0.0388)	(0.0464)	(0.0630)	(0.0598)	(0.103)
Net HH. Inc. €1801-2600	0.935**	0.869***	0.0656**	0.197***	-0.00238	-0.0363
	(0.0306)	(0.0218)	(0.0294)	(0.0509)	(0.0472)	(0.0825)
Educ. middle	0.826***	0.821***	0.00582	-0.138**	-0.224***	0.181*
	(0.0287)	(0.0231)	(0.0294)	(0.0557)	(0.0498)	(0.0985)
Educ. high	1.043	1.011	0.0321	-0.246***	-0.113**	-0.0379
	(0.0301)	(0.0268)	(0.0327)	(0.0508)	(0.0459)	(0.0928)
Health: good	1.323***	1.270***	0.0535	0.106	-0.137**	0.0520
	(0.0380)	(0.0350)	(0.0442)	(0.0660)	(0.0634)	(0.110)
Health: fair	1.720***	1.561***	0.159**	0.0918	0.129*	-0.176
	(0.0677)	(0.0568)	(0.0734)	(0.0785)	(0.0761)	(0.134)
Health: not good/poor	1.991***	1.956***	0.0346	0.0285	0.311***	-0.144
	(0.116)	(0.113)	(0.132)	(0.113)	(0.0998)	(0.175)
Constant	0.0138***	0.0113***	0.00254***	2.111***	2.342***	
	(0.000533)	(0.000501)	(0.000617)	(0.0865)	(0.0738)	
μ_1						-2.280***
						(0.168)
μ_2						-0.391**
						(0.164)
μ_3						1.344***
						(0.179)
μ_4						2.154***
						(0.195)
μ_5						3.412***
						(0.243)
Chi2 test joint equality (16df)	163.36*** ($p < 0.0001$)					
Chi2 test joint equality no cohorts (11df)	87.37*** ($p < 0.0001$)					
Baseline hazard	0.0763***	0.0834***				
	(0.00131)	(0.00154)				
Variance ind. effects	0.798***	0.497***				0.692***
	(0.0284)	(0.0176)				(0.109)
Variance seq. effects	0.0915***	0.0330***				0.00173
	(0.00673)	(0.00501)				(0.00506)
No. individuals			1,470			
No. probabilities			4,034			
Log-likelihood			-16,153.967			

[a] Estimates reported as hazard ratios.
Standard errors in parentheses; $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table F4: Gompertz model of *remaining* subjective survival with rounding – model estimated on all valid probabilities

	PB[a]	DHS[a]	Diff. PB - DHS	Error PB	Error DHS	Rounding PB
Coh. 1922-1931	5.906***	3.689***	2.218***	-0.133	0.438***	0.104
	(0.590)	(0.348)	(0.494)	(0.0835)	(0.0867)	(0.146)
Coh. 1932-41	3.278***	2.497***	0.781***	-0.0310	0.346***	0.149**
	(0.174)	(0.108)	(0.146)	(0.0354)	(0.0483)	(0.0691)
Coh. 1952-61	0.435***	0.640***	-0.205***	0.0285	-0.0554	0.0499
	(0.00942)	(0.0150)	(0.0150)	(0.0265)	(0.0395)	(0.0528)
Coh. 1962-71	0.210***	0.349***	-0.139***	-0.0405	-0.0359	-0.0416
	(0.00589)	(0.0104)	(0.00943)	(0.0287)	(0.0391)	(0.0568)
Coh. 1972-81	0.0744***	0.156***	-0.0816***	0.0657*	-0.122***	0.0411
	(0.00234)	(0.00460)	(0.00443)	(0.0355)	(0.0435)	(0.0675)
Coh. 1982-87	0.0262***	0.0673***	-0.0411***	-0.246***	-0.364***	0.173
	(0.00280)	(0.00512)	(0.00381)	(0.0761)	(0.0931)	(0.127)
Wave 2012	1.115***	1.066***	0.0489***	-0.0853***	0.0843***	-0.0346
	(0.0153)	(0.0123)	(0.0186)	(0.0231)	(0.0299)	(0.0326)
Female	0.911***	0.930***	-0.0193	0.0414**	0.0377	0.0550
	(0.0180)	(0.0181)	(0.0201)	(0.0204)	(0.0257)	(0.0390)
Net HH. Inc. $\leq$ €1150	1.154**	1.084**	0.0697	0.142***	0.151***	-0.00355
	(0.0663)	(0.0428)	(0.0625)	(0.0450)	(0.0521)	(0.0794)
Net HH. Inc. €1151-1800	0.903***	0.883***	0.0196	0.169***	0.0795**	-0.0498
	(0.0292)	(0.0250)	(0.0312)	(0.0307)	(0.0396)	(0.0562)
Net HH. Inc. €1801-2600	0.978	0.966*	0.0118	0.0946***	0.00389	-0.0397
	(0.0229)	(0.0197)	(0.0255)	(0.0242)	(0.0315)	(0.0454)
Educ. middle	0.819***	0.890***	-0.0712***	-0.0976***	-0.143***	0.0989*
	(0.0202)	(0.0212)	(0.0243)	(0.0270)	(0.0346)	(0.0516)
Educ. high	0.984	0.977	0.00742	-0.238***	-0.132***	-0.0210
	(0.0207)	(0.0210)	(0.0257)	(0.0257)	(0.0320)	(0.0498)
Health: good	1.225***	1.188***	0.0371	-0.0117	0.0634	-0.0276
	(0.0228)	(0.0229)	(0.0290)	(0.0297)	(0.0392)	(0.0578)
Health: fair	1.812***	1.549***	0.264***	0.0465	0.263***	-0.170**
	(0.0541)	(0.0448)	(0.0596)	(0.0366)	(0.0501)	(0.0706)
Health: not good/poor	2.135***	1.860***	0.276***	0.0215	0.305***	-0.151*
	(0.0956)	(0.0815)	(0.0997)	(0.0501)	(0.0679)	(0.0918)
Constant	0.0101***	0.0119***	-0.00176***	2.457***	2.193***	
	(0.000287)	(0.000325)	(0.000394)	(0.0393)	(0.0532)	
μ_1						-1.968***
						(0.0834)
μ_2						-0.396***
						(0.0809)
μ_3						1.205***
						(0.0840)
μ_4						1.919***
						(0.0922)
μ_5						2.977***
						(0.120)
Chi2 test joint equality (17df)	691.92 ($p < 0.0001$)					
Chi2 test joint equality no cohorts (11df)	60.86*** ($p < 0.0001$)					
Baseline hazard	0.0903***	0.0643***				
	(0.000660)	(0.00102)				
Variance ind. effects	0.846***	0.487***				0.354***
	(0.0223)	(0.0142)				(0.0361)
Variance seq. effects	0.112***	0.0340***				0.0116
	(0.00502)	(0.00408)				(0.00741)
No. individuals			2,323			
No. probabilities			16,540			
Log-likelihood			-40,715.570			

[a] Estimates reported as hazard ratios.
Standard errors in parentheses; ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Table F5: Correlation matrices of individual and question sequence effects for models of *remaining* lifetime

	Probabilities elicited in both surveys			All valid probabilities		
	PB	DHS	Round PB	PB	DHS	Round PB
a. Individual effects						
PB	1			1		
DHS	0.858***	1		0.827***	1	
Round PB	-0.0771	-0.0660	1	-0.183***	-0.0744*	1
	Probabilities elicited in both surveys			**All valid probabilities**		
	PB	DHS	Round PB	PB	DHS	Round PB
b. Sequence effects						
	PB	DHS	Round PB			
PB	1			1		
DHS	0.0460	1		0.429***	1	
Round PB	0.159	0.734	1	-0.447*	-0.646***	1

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$