Test-Retest Reliability of Subjective Survival Expectations[*]

J. de Bresser
University of Groningen[†]
Netspar[‡]

February 14, 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 mostly 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 the reliability of subjective well-being found by Krueger and Skade (2008). Correlations are weaker and averages differ more among respondents above the age of 65, which calls into question data quality for older respondents. 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 as indicated by the importance of individual effects. This unobserved heterogeneity is strongly correlated across surveys. Taken together this evidence supports the reliability of subjective survival expectations.

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

JEL-codes: D84; J14; C34

---

[*] I thank Netspar for financially supporting data collection. All errors are my own.

[†] University of Groningen, P.O. Box 72, 9700 AB Groningen, the Netherlands. Email: j.de.bresser@rug.nl.

[‡] Netspar, P.O. Box 90153, 5000 LE Tilburg, The Netherlands

1 Introduction

Expectations play an important role in economic models of intertemporal 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 evaluates for the first time the test-retest reliability of expectations reported by survey respondents. 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 come from a large household panel that is representative for the Dutch population: the CentERpanel. One survey, the Pensionbarometer (PB), allows respondents to report any integer probability between 0 and 100 percent. The other, the DNB Household Survey (DHS), restricts responses to an 11-point scale ranging from 0 to 10. Such 11-point scale limits the resolution at which respondents can report, forcing them to round their subjective probabilities. Nonetheless, it has been applied in several large scale household surveys, such as the Rand version of the HRS in the U.S., SHARE in Europe and the LISS panel in the Netherlands. However, with the exception of Bissonnette et al. (2011), researchers have interpreted the answers to 11-point scales as exact probabilities.

We evaluate the reliability of reported expectations in two ways. Firstly, we check whether the probabilities are consistent with each other by means of direct comparison. We compare probabilities reported by the same individuals for the same target ages one by one, taking into account that the different answer scales affect the resolution at which respondents report their expectations. As a result probabilities are rounded, which means that any reported probability is indicative of an interval within which the true probability falls (Manski and Molinari, 2010). Following the literature we distinguish between two different rounding schemes: common rounding for all reported survival probabilities, as in Manski and Molinari (2010), and conservative rounding where each individual probability may be rounded differently (see Bissonnette and de Bresser, 2014). We do not assume rounding is the same across the two response scales, since the DHS restricts respondents to round their probabilities to multiples of 10, 50 and 100 percent. The PB, on the other hand, allows more precise answers: we account for rounding to multiples of 1, 5, 10, 50 and 100 percent.

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 in one model. In doing so we allow for the possibility that discrepancies between individual probabilities cancel out when analyzed at a higher level of aggregation. We develop two versions of the model, one of which accounts for rounding while the other does not. For both versions we test whether the parameters that capture the correlation between background variables and the hazard of death are the same across surveys.

This paper fits in with the large literature on subjective expectations in general and survival expectations in particular (see Hurd, 2009, for an overview of research on subjective longevity). A rich body of literature 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). Plausible associations between subjective survival and background variables support 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). To the best of our knowledge the current paper is the first effort to evaluate the test-retest reliability of subjective expectations of any type.

We find that reported probabilities are reliable overall, but less so for older respondents above the age of 65. Our analysis of individual probabilities shows that test-retest correlations are between 0.5 and 0.7, which is comparable to the reliability of subjective well-being documented by Krueger and Skade (2008). Correlations are lower and the differences between the average reported probabilities are larger for the older target ages of 85 and 90, because those items were presented to older respondents. 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. A model in which all reported probabilities feature jointly shows 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 are important for both sets of probabilities and that unobserved heterogeneity is strongly correlated between datasets (correlation coefficients 0.8-0.9).

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. Section 6 concludes.

2 Survival questions in the Pensionbarometer and in the DNB Household Survey

Bot 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

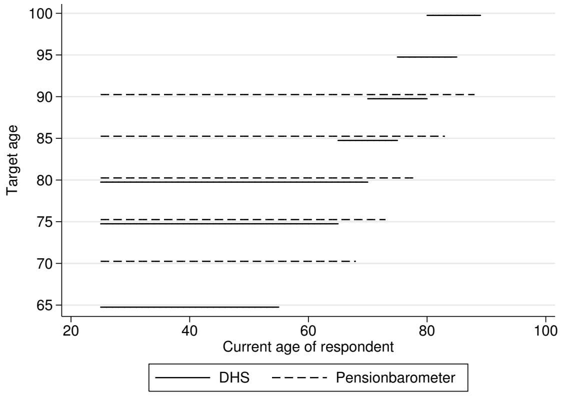

Figure 1: Age eligibility for survival questions in the DHS and in the Pensionbarometer

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 of 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. The notion that both questionnaires aim to measure the same expectations is plausible, since the period between questionnaires is short. For most individuals both surveys were conducted in June of 2011 and 2012. 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]

---

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

	N	Current age	Mean LT	PB		DHS		Rank corr.
				Mean	S. D.	Mean	S. D.	
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

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 are weakly decreasing with age and 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.

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

---

[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.

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 higher target ages the average probability from 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 a given aspect of well-being is usually measured by a single item in a questionnaire, but 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 percentage 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.

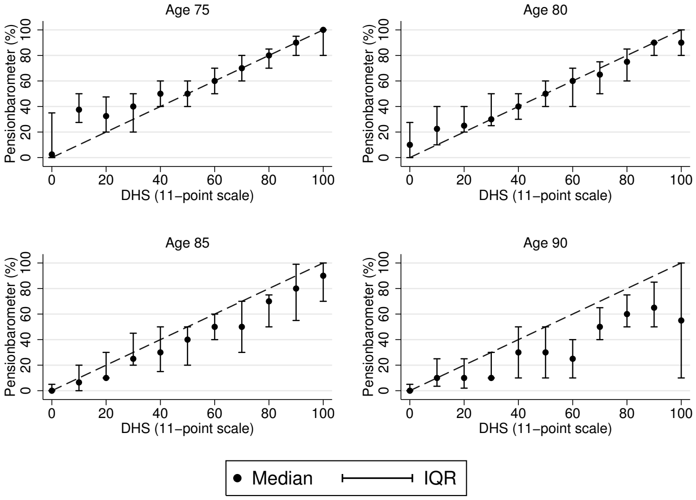

Figure 2: Medians and IQRs of survival probabilities in the Pensionbarometer conditional on responses to the corresponding DHS question

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 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 is 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

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.

$P^{PB} \in \left[P^{DHS} - 5, P^{DHS} + 5\right]$.[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).

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

---

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

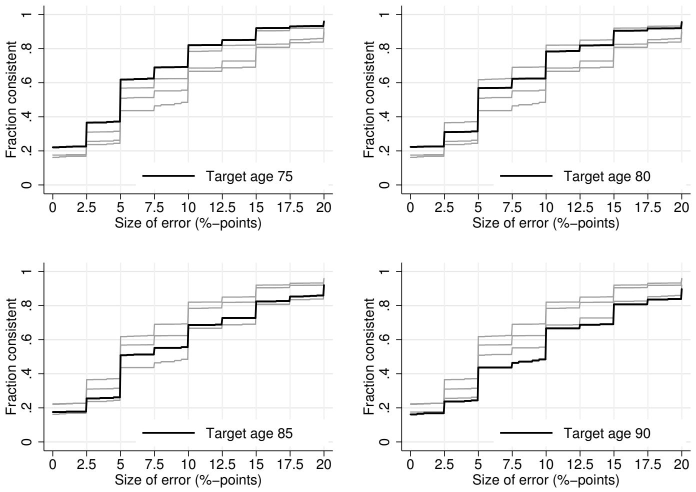

Figure 3: Fraction of probabilities that are consistent across PB and DHS while allowing for reporting noise

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 rises up to age 40, declines slightly between age 40 and 45 and declines further after age 65/70. 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 clearly related, it takes considerable rounding error for a majority of the cases in order to make the PB and DHS responses compatible with 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 Do both surveys yield similar survival curves?

4.1 Model without focal answers and 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. This parameterization of expectations implies that true probabilities on a scale from 0 to 100 are given by:

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

where q indexes questionnaires ($q \in \{PB, DHS\}$); $\gamma_{it}^{q} = \exp\left(\mathbf{x}_{it}'\boldsymbol{\beta}_1^{q} + \xi_i^{q} + \eta_{it}^{q}\right)$ 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_i^{q} and question sequence effects η_{it}^{q}. Distributional assumptions for these error components are given later. In the absence of unobserved heterogeneity the null hypothesis of interest is that $\boldsymbol{\beta}_1^{PB} = \boldsymbol{\beta}_1^{DHS}$ and $\alpha^{PB} = \alpha^{DHS}$, which implies that the two surveys measure the same underlying true probabilities.

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

$$P_{itk}^{*q} = S_{itk}^{q} + \varepsilon_{itk}^{q}$$

where $\varepsilon_{itk}^{q} \sim \mathcal{N}\left(0, \sigma_{it}^{2}\right)$, independent of all covariates and across thresholds, surveys, years and individuals. We model the variance of recall errors as $\ln\left(\sigma_{it}\right) = \mathbf{x}_{it}'\boldsymbol{\beta}_2^{q}$. 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 other error components. 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 $P^{*q}_{itk} \sim \mathcal{N}\left(S^q_{itk}, \sigma^2_{it}\right)$. As a given reported probability may result from different degrees of rounding, rounding is a latent construct and we average across the different rules 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):

$$f\left(P^q_{itk}|\mathbf{x}_{it}\right) = \sum_{r\in\Omega_{itk}} \Pr\left(R^q_{itk} = r|\mathbf{x}_{it}\right) \times \Pr\left(LB^r_{itk} \leq P^{*q}_{itk} < UB^r_{itk}|\mathbf{x}_{it}\right)$$

where $\Pr\left(LB^r_{itk} \leq P^{*q}_{itk} < UB^r_{itk}|\mathbf{x}_{it}\right)$ is given by

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

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. 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 C.

5.1 Model without rounding of reported probabilities

Estimation results of the model without rounding are presented in Table 3 (see section 4.1 for a detailed description of this model). The two leftmost columns present the effects of covariates on the baseline hazard as hazard ratios. These columns model PB and DHS probabilities respectively. The estimated effects of most covariates are both qualitatively and quantitatively very similar for the PB and the DHS. We find that older cohorts report a substantially higher hazard of death: relative to the baseline cohort 1942-1951 the hazard of the 1932-1941 cohort is a factor 2.47 higher according to the PB probabilities and a factor 2.14 higher according to the DHS. The hazard for the youngest cohort, born between 1982 and 1987, is only 4.7-5.1% of that of the baseline cohort. The hazard of death is 6.1-7.4% higher in 2012 compared to the baseline of 2011. Women report a lower hazard of death compared to men, but while this difference is significant for the PB probabilities (hazard ratio 91%), it is not significant for the DHS (hazard ratio 98%). We also find some disagreement between the PB and the DHS for the dummies corresponding to different levels of net household income. While the hazard from the PB probabilities is 12% higher for the middle income group of €1151-1800 per month than the baseline of more than €2600, the hazard ratio based on DHS probabilities is not significantly different from one. The education and subjective health dummies all show very similar patterns for the PB and the DHS: higher educated respondents and respondents who rate their current health more positively report lower hazards of death regardless of the set of probabilities used. For the education dummies the middle and high education groups report hazards that are on average 86-89% of that of the lowest education category. Furthermore, the average hazard of respondents who rate their health as “not good” or “poor” is 65-73% higher than that of respondents who rate their health as “excellent”.

Though both sets of probabilities yield associations between covariates and the baseline hazard that are quantitatively similar, the third column of Table 3 shows that standard errors are sufficiently small to make many nonetheless significantly different. This holds especially for the cohort effects, all of which are of the same sign and order of magnitude regardless of the set of probabilities modelled, yet all but that for the youngest cohort are significantly different. These differences between the cohort effects for the PB and for the DHS imply that older respondents report probabilities that are less reliable and more different between the surveys. The third column also highlights the different coefficients of the female-dummy and for the middle income group. Taken together, we reject joint equality of the hazard ratios for both datasets.

Columns 4 and 5 of Table 3 present estimates of the coefficients that govern the heteroskedasticity of the recall error, capturing variation in the extent to which reported probabilities fit the Gompertz distribution. The only factor that affects recall error similarly in both sets of probabilities is education: the highest education category of university graduates reports probabilities that are significantly less noisy compared to respondents who have not finished vocational training.

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: 0.075 for the PB and 0.079 for the DHS. The estimated variances of the individual effects indicate that expectations are persistent at the level of the individual for both datasets, though the PB shows more persistence than the DHS. Furthermore, the individual effects are strongly positively correlated with a correlation coefficient of 0.88. There is some additional clustering at the level of the survey wave, but

Table 3: Gompertz model of subjective survival without rounding

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

the variance of the sequence effects is much smaller than that of the individual effects.

Table C1 in Appendix C 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). Though this yields larger differences between the hazard ratios for cohort effects, the same general picture emerges. When analyzed jointly, hazard rates derived from both datasets show similar associations with covariates.

5.2 Model with rounding of reported probabilities

Estimates for the model that accounts for rounding, described in section 4.2 are reported in Table 4. Compared to Tabel 3 there is one new 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 log-likelihood function. In other words: estimation strongly indicates that all probabilities in DHS are rounded to multiples of 10.

Accounting for rounding does not change the conclusion that similar relationships between the hazard of death and covariates emerge for the PB and the DHS. Cohort dummies all enter the model with similar hazard ratios for the PB and DHS, though the difference between the two is statistically significant for all but the 1972-1982 cohort. The differences between cohort effects are largest for the oldest cohorts, which confirms again that reliability is lower for older respondents. The hazard of death was 7-9% higher in the second wave, the difference between the PB and DHS is not statistically significant. Moreover, when we account for rounding women have a significantly lower hazard of death according to both datasets, while that is only the case for the PB when we do not model rounding. The difference relative to men becomes much larger in the extended model: their hazard is only 75% of that of men according to the PB and 83% according to the DHS (the difference between these hazard ratios is significant). For income we also find that results of the second model differ from those of the baseline: according to both surveys individuals from the second highest income group have a significantly lower hazard relative to the highest income group (hazard ratio 0.94 for PB and 0.87 for DHS, difference significant). The difference in expectations between the highest and the lowest education groups is no longer significant once we account for rounding, but the middle education group still has a lower hazard compared to the lowest group. For current health the model with rounding corroborates the large differences found in the model without rounding: the hazard for people in poor health is almost twice that of individuals in excellent health. However, the difference between the hazard ratios for the "fair" category according to the PB and the DHS is now statistically significant at the 5 percent level, while it wasn't before.

In summary we conclude that accounting for rounding does not affect the extent to which both datasets yields similar results regarding the relationships between the hazard of death and socio-demographic covariates. We reject equality between the estimated hazard ratios slightly more often than was the case in the baseline model. However, qualitatively both sets of probabilities still yield similar patterns. These patterns are mostly similar to those found in the model without rounding. Exceptions are a larger difference in hazards between men and women and between some income groups and a smaller difference between the highest and lowest education groups in the model with rounding compared with the baseline model.

The fourth and fifth column in Table 4 present estimates for the heteroskedasticity equations for the recall error in the PB and DHS respectively. We note that the variance of the errors is significantly lower among higher education groups, as was the case in the model without rounding. The coefficients of the rounding equation for the PB, shown in the final column, are not estimated precisely. Only the middle education dummy is marginally significant, indicating that those in that category round more roughly than do their poorly

Table 4: Gompertz model of subjective survival with rounding

	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 5: Correlations of individuals and sequence effects

a. Individual effects			
	PB	DHS	Rounding PB
PB	1		
DHS	0.858***	1	
Rounding PB	-0.0771	-0.0660	1

b. Sequence effects			
	PB	DHS	Rounding PB
PB	1		
DHS	0.0460	1	
Rounding PB	0.159	0.734	1

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

educated peers. Though we do not observe clear evidence for variation in rounding with socio-demographic variables, we do estimate the thresholds between different levels of rounding quite precisely. In particular, none of the 95 percent confidence intervals overlap, which indicates that we successfully identify the fractions of individuals that use different rounding rules. While the model demonstrates that all DHS probabilities are rounded to multiples of 10, it does find different degrees of rounding for the PB.

The baseline hazard is similar across the PB and the DHS, and with a value of 0.08 it is also 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, but more so for the PB than for the DHS. Question sequence effects are also significant, but much smaller in magnitude. Rounding too is persistent at the level of the individual for the PB, but we find no additional persistence across probabilities in a given sequence of questions. 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.

6 Conclusion

There is a growing body of research that 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. Furthermore, we formulate a model in which we analyze reported probabilities jointly so we can 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 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). Especially for men we find lower correlations for older target ages of 85 and 90, and for both men and women the average reported probability is around 10 %-points lower in the PB than in the DHS for those target ages. Further analysis reveals that this is due to the effect of the current age of respondents: older respondents report less reliable probabilities. 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 dummies that correspond to the oldest age cohorts, for which we find that the hazard of death is higher in the PB than in the DHS. 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. Furthermore, incorporating rounding in the model does not reduce differences between the estimates from both datasets.

Taking all results together we conclude that subjective survival expectations are as reliable as subjective well-being data which are frequently analyzed by economists. 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. Netspar Discussion Paper.

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. 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.

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 based on all valid observations

Table C1: Gompertz model of subjective survival without rounding

	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 C2: Gompertz model of subjective survival with rounding

	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 C3: Correlation matrices of individual and question sequence effects estimated from all valid probabilities

a. Individual effects

	PB	DHS	Round PB
PB	1		
DHS	0.827***	1	
Round PB	-0.183***	-0.0744*	1

b. Sequence effects

	PB	DHS	Round PB
PB	1		
DHS	0.429***	1	
Round PB	-0.447*	-0.646***	1

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$