Chaire de recherche Industrielle Alliance
sur les enjeux économiques
des changements démographiques

CEDIA.ca

# 15-03

# ELICITING SUBJECTIVE SURVIVAL CURVES: LESSONS FROM PARTIAL IDENTIFICATION

CAHIER DE RECHERCHE
WORKING PAPER

Luc Bissonnette and Jochem de Bresser

Février / February 2015

Université Laval
Faculté des sciences sociales

ESG UQÀM

La Chaire de recherche Industrielle Alliance sur les enjeux économiques des changements démographiques est une chaire bi-institutionnelle qui s'appuie sur un partenariat avec les organisations suivantes :

**- Centre interuniversitaire de recherche en analyse des organisations (CIRANO)**
**- Industrielle Alliance, Assurance et services financiers inc.**
**- Régie des rentes du Québec**



Dépôt légal : 1er trimestre 2015
ISSN 2368-7207

# Eliciting Subjective Survival Curves: Lessons from Partial Identification*


L. Bissonnette
Université Laval[†]

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


February 17, 2015


**Abstract**

When analyzing subjective expectations, researchers commonly apply (non-)parametric approximations to point identify beliefs. We propose a new take on this type of data that does not impose a functional form on expectations. Using the widely researched example of subjective survival expectations, we construct bounds for subjective survival curves. These bounds allow us to partially identify subjective life expectancy. We show that the informativeness of the bounds depends on our willingness to interpolate beliefs between data points. If we do not smooth between the elicited points on the survival functions, the bounds are too wide for useful inference. However, if we do interpolate and allow for a limited amount of rounding, the resulting bounds are narrow enough to show variation in life expectancy with age and self-reported health, the strongest predictors in point identified models. Finally, we match the subjective data to life tables. While analysis that point identifies life expectancy, either parametrically or non-parametrically, rejects consistency of expectations with actuarial forecasts for women, the bounds show that allowing for rounding renders the subjective data consistent with forecasts on average.



**Key words:** Subjective expectations, life expectancy, partial identification
**JEL Codes:** C81, D84, I19


---


*We thank Netspar for financially supporting the data collection. Moreover, we thank Arthur van Soest, Frederic Vermeulen, Martin Salm, Marcel Das, Pierre-Carl Michaud, Rob Alessie, Charles Bellemare and participants at the 2014 MiSoC Workshop on Subjective Expectations and Probabilities in Economics and Psychology for their insightful comments.



[†]Departement d'Économie, Université Laval, 1025, avenue des Sciences-Humaines, Québec (Québec), Canada, G1V 0A6. Email: Luc.Bissonnette@ecn.ulaval.ca.

[‡]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

Human beings take into account the future consequences of their behavior. In economics this notion is formalized through models in which rational agents fully grasp the inter-temporal effects of their decisions, integrating the present with expectations of what is to come. For instance, when deciding how much to save for retirement, one combines preferences over current and future consumption with ideas of how long this future might be: with beliefs about one's longevity (e.g. Hurd et al., 2004). While we can expect that different individuals hold different beliefs regarding their longevity, it is usually impossible to identify separately the role of preferences and expectations in decision making unless data on subjective expectations are available (see Manski, 2004, for a review). The common solution is to assume that the formation of expectations is rational, in the sense that they coincide with a distribution observed by the researcher. In the specific context of mortality, beliefs are usually approximated by actuarial tables (taking into account birth cohort and sex). However, equating expectations to life tables misses important dimensions of heterogeneity (e.g., a smoker can expect to live a shorter life). Moreover, individuals' subjective beliefs may not correspond to the official life tables in the aggregate (Peracchi and Perotti, 2011). Previous research shows that especially women on average expect to die earlier than predicted by actuarial tables (see Perozek, 2008, for the US). For these reasons researchers have started to analyze expectations directly by means of survey questions eliciting subjective beliefs (see Hurd, 2009, for a recent review). These elicited subjective expectations have been used to estimate choice models (e.g., Van der Klaauw and Wolpin, 2008; Bellemare et al., 2008; and Van der Klaauw, 2012, provide examples of how expectations elicited in surveys can be combined with observed decisions to estimate dynamic structural models). There is a large literature that documents heterogeneity in subjective longevity and establishes the predictive validity of subjective longevity for survival at the individual level (e.g. Hurd and McGarry, 1995, 2002; Smith et al., 2001; Bissonnette et al., 2011, for the US; and Kutlu and Kalwij, 2012, for the Netherlands).

Analyses of subjective expectations often rely on approximations to bridge the gap between the information contained in the data and the fully defined distributions that economists are interested in. We propose a new method for the analysis of subjective expectations of a continuous outcome such as next year's wages or life expectancy. For continuous variables the data contain points on the cumulative distribution function over future events (in our application respondents report the probability of survival past five age thresholds). Such data have been analyzed in ways that point-identify expectations, either by assuming some underlying parametric distribution, or by approximating expectations flexibly by means of splines. Examples of the parametric approach are Dominitz and Manski (1997) and Dominitz (1998) on income expectations; Dominitz and Manski (2006) on pension benefit expectations and Perozek (2008), among others, on mortality. The non-parametric spline technique was introduced by Bellemare et al. (2012), who also show that it is robust to rounding of the reported probabilities. We compare the point identified approaches with one that does not impose any restrictions on expectations. This novel method derives bounds on life expectancy that take into account that surveys only collect information on some points on the CDF. Instead of constructing a unique CDF based on these observed points on the distribution, we propose to consider all the distributions that are coherent with the elicited information. We show that the resulting non-parametric bounds can easily be generalized to allow for rounding of the reported probabilities. This approach of bounding distributions rather than fitting them is inspired by numerous papers by Charles Manski, a leader in the studies of partial identification of distributions under weak assumptions (see, for example, Manski, 2003). Our paper is close in spirit to two papers he co-authored. Engelberg et al. (2009) explore the idea of bounding measures of the central tendency using probabilistic information that is similar to the data we will use in this paper. Manski and Molinari (2010) study rounding systematically, but in the context of expectations regarding binary outcomes that are fully described by a single subjective probability. We depart from these studies in two ways. Firstly, we are explicitly interested in bounding an entire distributional curve, since the probability to survive to any given age can be a value of interest (for instance, to model inter-temporal behavior in a dynamic model). Secondly, we study the implications of rounding in the context of expectations of a continuous variable more thoroughly than is done in Engelberg et al. (2009). Our analysis provides various results, illustrating what can be learned about expectations under different sets of assumptions.

Using a sample of Dutch adults, we show that the bounds on life expectancy are wide if we do not smooth expectations between data points: they are 11 years wide on average. Allowing for rounding increases the width of the bounds, making them even less informative. However, we can narrow the bounds substantially if we restrict the set of admissible distributions. We propose to do so by simultaneously interpolating beliefs between data points and allowing for rounding. Such interpolation using linear or cubic splines does not impose a parametric form, but does assume subjective survival functions are continuous and piecewise linear (linear splines) or continuous and smooth (cubic spline) (Bellemare et al., 2012). Furthermore, interpolation allows us to disentangle the effects of rounding of reported probabilities and the limitations inherent in modeling expectations using small number of probabilities. Interpolation reduces the average width of the bounds to 3 years if we allow for rounding but assume that each probability reported by a respondent is rounded to the same extent. The average width is 7 years if we allow each probability to be rounded differently, still an improvement compared to bounds without smoothing or rounding.

To illustrate the importance of our analysis, we apply the econometric framework for the analysis of interval-censored data to investigate differences in life expectancy across socioeconomic groups. Tamer (2010) provides a summary of that literature and Imbens and Manski (2004) and Beresteanu and Molinari (2008) build the theory of best linear prediction of interval-censored dependent variables. If we are unwilling to impose any restrictions on expectations beyond what is given in the data, the bounds on life expectancy are too wide to allow for useful inference. In particular, partially identified models do not corroborate the strong relationship between life expectancy and the covariates *age* and *self-reported health* found in point identified linear models. Under the common rounding scheme the bounds derived using interpolated survival functions are sufficiently tight to corroborate the correlations between life expectancy and the covariates *age* and *health*.

Finally, we match our sets of bounds on life expectancy with life tables produced by Statistics Netherlands. The extent to which expectations correspond to actuarial forecasts on average is relevant, because economists often use life tables as substitutes for subjective expectations thanks to the easy availability of actuarial figures (Peracchi and Perotti, 2011). Our results suggest that some conclusions with regard to expectations disappear once we let go of point identification. If we point identify life expectancy, either through parametric restrictions on beliefs or through linear or cubic splines, we find that the expectations of men correspond closely to actuarial tables on average, but that women expect to live much shorter than the life tables suggest. The same discrepancy has been documented for American women by Perozek (2008). However, if we do not interpolate expectations, we cannot reject that beliefs are on average consistent with actuarial forecasts. This conclusion emerges even more strongly when we allow for rounding and remains unchanged if we restrict expectations to exhibit a weakly increasing hazard of death. If we do smooth beliefs, we cannot reject consistency of women's expectations with life tables if we allow each individual probability to be rounded to the maximum extent. If we impose that all probabilities reported by a respondent follow a common rounding scheme, we restore the result that women on average expect to die younger than life tables suggest for limited age-ranges around age 30 and 60.

These findings offer some insights into the nature of subjective probabilities that may be relevant for a broader group of researchers than those interested in subjective survival. Firstly, point identification of expectations by parametric distributions yields almost identical results to approximation by (linear or cubic) splines, at least if a sufficient number of subjective probabilities are available. We observe five probabilities and the correlations between the calculated life expectancies are above 0.98. If we do not mind interpolating expectations between reported probabilities and if we are comfortable to ignore rounding, results are robust with respect to the specific form used. Secondly, the coarseness of the grid on which expectations are elicited limits the informativeness of the data much more than the extent of rounding in the reported probabilities. In particular, even though we have a relatively rich set of five probabilities that we use to approximate expectations, we cannot learn much about subjective life expectancies if we do not smooth expectations between those points. Unfortunately, that is the case even if we do not allow for any rounding. Thirdly, if we do interpolate expectations, rounding does not necessarily make the data uninformative. However, the type of rounding matters: a common rounding scheme as proposed in Manski and Molinari (2010) leads to much more informative bounds on expectations than does a general "worst case"-scheme in which individual probabilities are rounded to the maximum extent. From that perspective it is reassuring that De Bresser and Van Soest (2013b) find that individuals tend to stick to the same rounding rule, suggesting that a single rounding rule may be an appropriate assumption.

The structure of the paper is as follows. Section 2 introduces the type of data that we use and explains the parametric and non-parametric methods that we apply to approximate expectations. The data are described in section 3 and section 4 presents our results. Section 5 concludes.

# 2 Methods

## 2.1 Survival questions

The subjective longevity questions that we analyze are similar to those found in the Health and Retirement Study (HRS). The HRS is a large household panel that is representative for the U.S. population of age 50 and older. It was the first survey to include questions on subjective longevity and, like our data, the HRS items also ask for the probability of surviving past certain age thresholds. However, in contrast to the HRS we consider questions that refer to a maximum of five thresholds, depending on the age of the respondent at the time of the questionnaire. The items are phrased as follows:

> Please indicate on a scale from 0 to 100 how likely you think it is that you live to:
>
> 1. *[if age $<$ 69]* age 70
> 2. *[if age $<$ 74]* age 75
>
> ...
>
> 5. *[if age $<$ 89]* age 90

Age-eligibility requires a respondent to be at least 2 years younger than a particular age threshold in order for that question to be presented. One example of possible answers for a respondent who is 68 or younger is given in Table 1. Depending on the estimation method and the rounding rule we will interpret these probabilities as perturbed probabilities from a distribution in a pre-determined family, as exact points on the subjective survival function that does not belong to a given family of distributions, or as rounded approximations to the survival function (see section 2.5.1 for an explanation of the different forms of rounding that we distinguish).

The fact that we observe five probabilities for many respondents means that our data on expectations is relatively rich compared to the HRS, which only contains two questions

**Table 1:** Hypothetical data on subjective survival expectations

| Event | Prob. (%) |
|---|---|
| Survive to age 70 | 90 |
| Survive to age 75 | 70 |
| Survive to age 80 | 45 |
| Survive to age 85 | 30 |
| Survive to age 90 | 15 |

on subjective longevity. This facilitates the estimation of parametric models of survival and increases the informativeness of the bounds that we derive.

## 2.2 Parametric survival functions

One commonly used approach when analyzing subjective data of the type described in the previous section is to fit parametric, observation-specific survival functions by non-linear least squares. This methodology was introduced by Dominitz and Manski (1997) in the context of expectations about future income and has been applied to subjective survival probabilities by Perozek (2008). The latter takes two popular functional forms for survival functions, the Gompertz and Weibull distributions, and obtains parameters for each individual by solving:

$$\min_{\theta_i^1,\theta_i^2} \sum_{t \in T_i} \left[ P_{it} - S\left(t, age_i; \theta_i^1, \theta_i^2\right)\right]^2 \tag{1}$$

Where $P_{it}$ is the subjective probability of survival past age $t$ reported by individual $i$; $T_i$ is the (age-dependent) set of all age thresholds $t$ for which the questions are asked (in the context of our data $T_i = \{70, 75, 80, 85, 90\}$ for respondents aged 68 or younger); $S(.)$ is the survival function that follows from the parametric form imposed on expectations and $\theta_i^1$ and $\theta_i^2$ are the parameters of $S(.)$ (these parameters are specific to individual $i$). Estimation of subjective survival curves by non-linear least squares assumes that the subjective probabilities of individuals are equal to the parametric probabilities $S(.)$, perturbed by mean zero IID errors. Figure 1 illustrates this method for the hypothetical data given in Table 1.

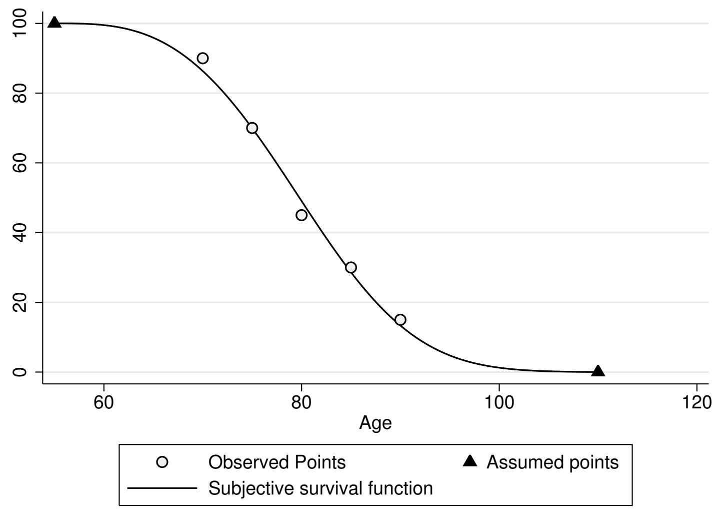


**Figure 1:** Point identification of expectations using a parametric, Weibull, survival function

In order to enforce that expectations are such that the probability of survival tends to zero for extremely old ages, we apply the technique introduced by Perozek (2008) and construct an additional probability for survival to age 98. This extra probability consists of the product of the subjective probability of survival past age 90 and the actuarial estimate of the likelihood of surviving to age 98, conditional on survival to age 90. For the hypothetical data from Table 1, this auxiliary probability would be equal to $P\left(90|age_i\right) \times P\left(98|90, age_i\right) = 15 \times P\left(98|90, age_i\right)$.[1] However, given that we observe five probabilities in the data, our estimates are not sensitive to this additional probability. After estimating the individual-specific parameters, we can easily calculate remaining life expectancy from the parametric survival curves.

[1]We only construct the probability of living past age 90 for the part of the sample that is at least 50 years old, because Statistics Netherlands does not yet offer forecasts of survival probabilities to that age for younger cohorts.

## 2.3 Non-parametric survival functions

An alternative approach that also allows for the point identification of expectations is to trace the points on the subjective survival function that are given in the data, without assuming that those probabilities stem from a parametric model. This method constructs a flexible (piecewise) function that passes exactly through the reported probabilities; it does not invoke implicit error terms as does the non-linear least squares method explained in the previous subsection. We tried approximating expectations with linear and cubic splines; the latter are explained by Bellemare et al. (2012). In the remainder of the paper we only report results from cubic splines, since Bellemare et al. (2012) show that those functions are best able to approximate a wide range of parametric models. Moreover, in the specific context of human mortality it seems unlikely that individuals have piecewise linear expectations, especially in the tails of the distribution.[2] However, cubic splines lead to complications when constructing intervals for life expectancy under the general rounding scheme described in section 2.5: the upper and/or lower bound cross the point estimate for 22% of the sample. Therefore, it is reassuring to find that the results from linear spline functions, for which the bounds never cross each other, are almost identical to those reported below. Figure 2 illustrates the spline method: it shows the fitted survival function using cubic splines for the example data given in Table 1.

Figure 2 also illustrates that the probabilities reported in the data, shown as circles in the figure, do not suffice to characterize expectations completely if we follow the spline approach. Only if the respondent is certain to survive past the youngest age threshold for which the respondent is eligible, $P_{it_{i1}} = 100$, and certain to die before the age of 90, $P_{i90} = 0$, do the data tie down expectations. This is an important consideration since 84 percent of respondents indicate less than complete certainty about their survival past the the first age threshold

[2]We expect the survival function to be concave for ages close to the current age of the respondent and convex for the oldest ages.

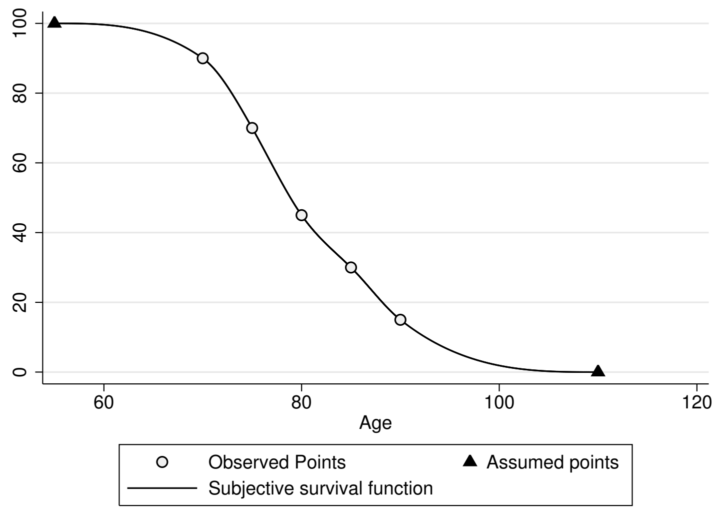


**Figure 2:** Point identification of expectations using cubic spline interpolation to construct the survival function

that was presented to them, so that these respondents require an imposed lower bound on the age of death. An even larger fraction of 87 percent indicate that the probability of living past the age of 90 is more than zero. The auxiliary probabilities that we construct are given by the triangles in Figure 2. For the minimum age of death we take the current age of the respondent (age 55 for the example in Figure 2). It is less straightforward to come up with a credible maximum age of death. We experimented with the ages 120, 110 and 100; in Figure 2 and the remainder of this paper we impose 110 as the maximum attainable age. Note that the influence of this maximum lifespan on the calculated life expectancies depends on the level of the final reported probability for age 90: larger probabilities will result in more sensitive life expectancies. Moreover, cubic spline interpolation tends to be more robust than linear splines with respect to the age of certain death, since the interpolated functions are usually convex in the final interval (see Figure 2). Results using the alternative maximum ages of 120 and 100 are similar to those reported in the text and are available on request.

## 2.4 Non-parametric bounds on life expectancy

Approximation of beliefs by non-linear least squares or spline functions is convenient, because both methods give us point estimates of the life expectancy of each individual in the sample (and of any other interesting moments of the distribution that characterizes expectations). However, neither road to point identification is free of potholes. For the non-linear least squares approach, there is some tension between first invoking an estimation procedure for the parametric distributions, relying explicitly on the idea of some kind of reporting error, and then analyzing the estimated distributions as if they are identically equal to expectations. Such a two-step approach does not take into account the effect of reporting errors on the subsequent analysis. Another disadvantage of the non-linear least squares method is that it requires one to impose some parametric form on expectations. Though the Gompertz and Weibull distributions might adequately model human survival, there is no guarantee that such descriptive accuracy for objective survival translates into descriptive adequacy for subjective expectations. Neither of these objections applies to flexible spline approximation, since the constructed survival functions follow the data exactly and impose no parametric restrictions on expectations. However, spline interpolation does smooth expectations between the reported probabilities in the data. Though smooth functions may reflect beliefs accurately, we want to investigate how much we can learn about expectations without relying on any parametric assumptions or interpolation.

In order to derive bounds on remaining life expectancy, we start from the same point as for the spline interpolation approach and interpret the reported probabilities as points on the subjective survival function. However, we do not assume any particular form of expectations between those points. The implication of this is that our data identifies boxes within which the subjective survival curve lies, but contains no information on the location of the curve within the boxes. The grey area in Figure 3 is the admissible set for survival expectations under the assumption that reported probabilities lie exactly on the subjective survival curve.

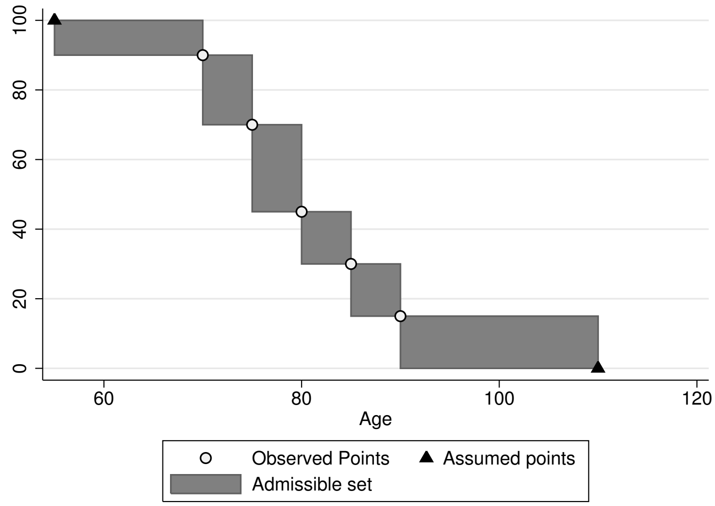


**Figure 3:** Admissible set for the survival curve (no smoothing between data points and no rounding)

Note that any function is allowed, even a step function, as long as it passes through the points given in the data. Though this makes the procedure very flexible, it does not by itself allow us to calculate bounds on life expectancy for a large majority of cases: in order to construct the bounds on life expectancy we need to impose a minimal and maximal age of death. As in the case of cubic spline interpolation explained in section 2.3, we assume death occurs between the current age of the respondent and age 110 (results with maximum ages equal to 100 and 120 are similar and available on request). Given the reported subjective probabilities and our constructed age-range across which death may occur, we simply trace the bottom edges of the boxes in Figure 3 to obtain the most pessimistic survival curve. Likewise, the upper edges yield the most optimistic curve that is consistent with the subjective data. Having bounded the survival function, we calculate the lower bound for life expectancy from the most pessimistic survival function and the upper bound from the most optimistic function.

## 2.5 Rounding

### 2.5.1 Common and general rounding schemes

Another methodological consideration in our analysis of subjective longevity is the effect of rounding of subjective probabilities on our inference. Rounding means that we cannot interpret the reported probabilities as points that are exactly on the subjective survival curve. Instead, rounded probabilities are informative of intervals within which the true subjective probabilities fall. For instance, a probability equal to 20 percent that is rounded to a multiple of five indicates that the true probability is in the interval $[17.5, 22.5]$, and the same probability rounded to a multiple of ten means that the true probability is in $[15, 25]$. This example shows that different rounding rules are observationally equivalent when we consider one single probability. Therefore, we apply the strategy proposed in Manski and Molinari (2010) to infer the degree of rounding from a set of related probability questions under the assumption that they result from the same rounding rule.[3] That is, we assume that the answers to all subjective longevity questions from a given respondent are rounded similarly and select the most conservative rounding rule that is consistent with all probabilities reported by an individual. Following Manski and Molinari (2010), we allow for the following types of rounding:

- All probabilities equal to 0 or 100: $P_{ik} \in [P_{LB}, P_{UB}] = [\max\{0, P_{ik} - 50\}, \min\{P_{ik} + 50, 100\}]$
- All probabilities equal to 0, 50 or 100: $P_{ik} \in [P_{LB}, P_{UB}] = [\max\{0, P_{ik} - 25\}, \min\{P_{ik} + 25, 100\}]$
- All probabilities multiples of 10: $P_{ik} \in [P_{LB}, P_{UB}] = [\max\{0, P_{ik} - 5\}, \min\{P_{ik} + 5, 100\}]$

[3]There is some evidence that supports the assumption of a common rounding rule for different questions that describe a single probability distribution is plausible. For instance, De Bresser and Van Soest (2013b) estimate a model that fits subjective distributions for the replacement rate of income at retirement from 6 probabilistic questions that are comparable to the survival questions analyzed in this paper. They allow different rounding for questions within a single sequence, but include survey and individual effects to describe persistence in rounding behavior. Together these effects account for 55-60 percent of total error variance, the remaining 40-45 percent being idiosyncratic variation that differs between questions in a sequence.

- All probabilities multiples of 5: $P_{ik} \in [P_{LB}, P_{UB}] = [\max\{0, P_{ik} - 2.5\}, \min\{P_{ik} + 2.5, 100\}]$

- Some probabilities in $\{1, 2, 3, 4\}$ or $\{96, 97, 98, 99\}$:

  - $P_{ik} \in \{0, 1, 2, 3, 4, 5\} \cup \{95, 96, 97, 98, 99, 100\}$: $P_{ik} \in [P_{LB}, P_{UB}] = P_{ik}$

  - $P_{ik} \in [6, 94]$: $P_{ik} \in [P_{LB}, P_{UB}] = [\max\{0, P_{ik} - 2.5\}, \min\{P_{ik} + 2.5, 100\}]$

- Not in any of the categories above: $P_{ik} \in [P_{LB}, P_{UB}] = P_{ik}$

Figure 4 illustrates how rounding affects the information contained in reported probabilities for the hypothetical data from Table 1. All probabilities in Table 1 are multiples of 5 and two probabilities are not multiples of 10, so according to the Manski and Molinari rounding scheme probabilities are rounded to multiples of 5. The thick bars around the reported probabilities in Figure 4 are the intervals within which subjective probabilities lie when we allow for rounding, assuming that all probabilities are rounded to the same extent. The width of those intervals is equal to 5 percentage points for all probabilities. We refer to this model for rounding as the *common* rounding scheme.

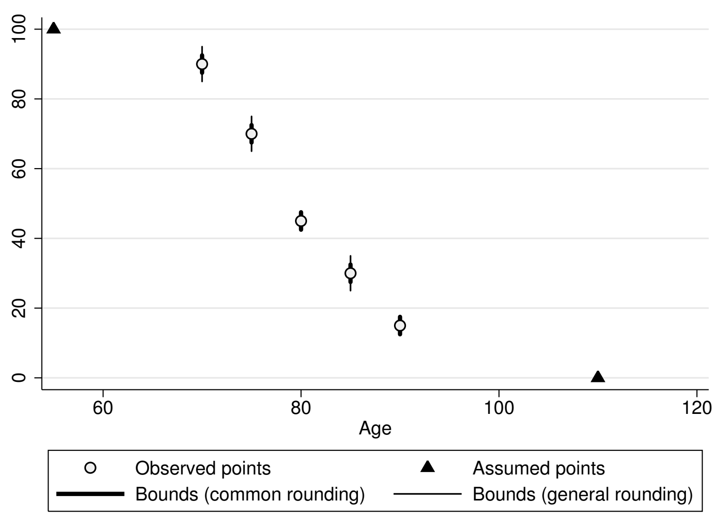


**Figure 4:** Bounds on true probabilities under different rounding rules

An alternative approach allows the extent of rounding to vary across the probabilities that describe an individual's expectations. Following this logic, we assume that reported probabilities are all rounded to the crudest extent possible. We allow for rounding to multiples of 100, 50, 25, 10 and 5. That is, a reported probability of 100 is interpreted as evidence that the true probability lies in the interval $[50, 100]$ and a reported 35 implies the interval $[32.5, 37.5]$, regardless of the other probabilities reported by that respondent. Probabilities that can only result from rounding to a multiple of 1 are interpreted as indicative of an interval with width 5 (we do not interpret any probabilities as exact representations of underlying expectations). Because this scheme does not restrict the probabilities reported by an individual to be rounded to the same extent, we call it the *general* rounding scheme. The thin bars in Figure 4 illustrate this conservative rounding rule for the example data from Table 1. Note that the general rounding rule does not affect the admissible set for those age brackets at which the reported probability was rounded to a multiple of 5 but not of 10, since the common rounding rule assumes that all probabilities are rounded to multiples of 5. In this case general rounding only increases the width of the intervals for the true probabilities at those thresholds for which the reported probability is a multiple of 10.

The general rounding scheme leads to broad intervals for some probabilities, especially for those equal to 0, 50 and 100. However, imposing monotonicity on the true, partially identified, probabilities helps to narrow down the bounds on reported probabilities to informative levels. Monotonicity implies that the lower bound on the true probability for any threshold can never be below the lowest possible probability at the following threshold. Similarly, the upper bound at any threshold can never be above that at the previous threshold. For instance, if we observe a probability equal to 50% for the first age threshold of 70, the general rounding scheme interprets that probability as indicative of an interval equal to $[25, 75]$ for the corresponding true probability. However, if the reported probability for age 75 is 40%, we know that the true probability for age 70 cannot be smaller than the lower bound of 35%

(since that is the lower bound on the probability for age 75).

### 2.5.2 Admissible sets for survival functions in the presence of rounding

According to the models of rounding explained in the previous section, probabilities reported by an individual respondent are either all generated by a common rounding rule, or they are rounded one by one to the maximum extent. We construct admissible sets for the corresponding survival function for both rounding models by tracing the upper and lower bounds of the intervals for the true, unobserved, probabilities.

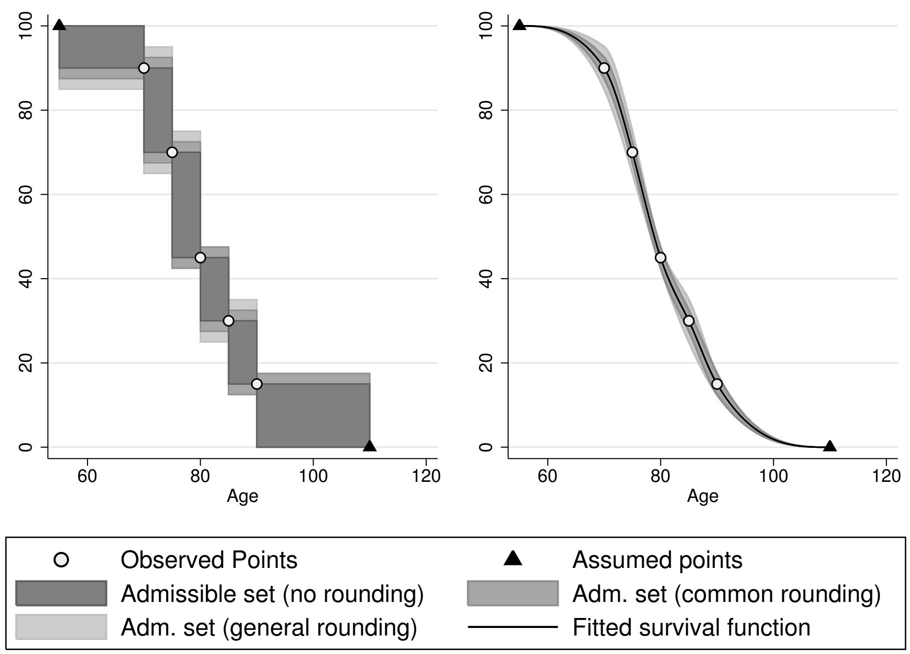


**Figure 5:** Admissible sets for the survival function under different rounding rules. Left panel: no smoothing; right panel: smoothing between reported probabilities.

Figure 5 presents the admissible sets that contain the true survival function under both rounding rules for the example data from Table 1. The left panel does not smooth expectations between the reported probabilities, illustrating how the admissible set changes due to rounding relative to the scenario in which we do not make any assumptions beyond exact reporting of true probabilities (see Figure 3). The right panel of Figure 5 shows admissible sets obtained by simultaneously allowing for rounding and interpolating expectations between age thresholds by means of cubic splines as in Figure 2. In both panels the lightest grey band traces the upper and lower bounds on probabilities constructed using the more general and conservative rule that probabilities are rounded individually. The darker area assumes that all probabilities reported by a respondent are rounded to the same extent. The width of the identified region differs between rounding rules only at those thresholds for which the reported probability is not only a multiple of five, but also a multiple of 10 (the ages 70, 75 and 85). The survival function that yields the highest life expectancy that is consistent with the data and the rounding rule is that which traces the upper bound of the identified region, while the lowest life expectancy is obtained by tracing the lower bound.

Though uncertainty does not receive much attention in the literature on subjective longevity, our method of partially identifying survival curves does allow one to construct bounds for measures of dispersion. Two measures that are used to capture subjective uncertainty in other domains are the standard deviation and inter-quartile range (IQR) of the fitted distribution. Bounds on the IQR can easily be computed by locating the ages at which the lower and upper bound on the subjective survival function are equal to 75% and 25%. However, even if we do not allow for rounding the bounds for the IQR are sensitive to the maximum age imposed if the reported probability of living past the age of 90 is 25% or more, which is the case for 45% of our sample. It is more difficult to derive bounds on the standard deviation of a subjective distribution for a given identified region. Moreover, the survival curve that maximizes the standard deviation places as much probability mass as possible in the tails of the distribution, making the bounds sensitive to the choice of maximum lifespan.

We tried to reduce the size of admissible sets without interpolation between reported probabilities by imposing that survival curves are continuous and that the hazard of dying increases weakly with age. However, 65% of respondents report probabilities that are incompatible with the assumption of an increasing hazard if we do not allow for rounding. If we do allow for rounding, the restrictions on expectations yield bounds on life expectancy that are not much narrower on average than those derived without rounding or restrictions on expectations. Results are available on request.

# 3 Data quality and descriptives

We use the 2011 wave of the yearly *Pensioenbarometer*, a survey administered to the respondents of the CentERpanel. Data collection was financed by Netspar, Network for Studies on Pensions, Aging and Retirement. The CentERpanel is administrated by CentERData and is representative of the Dutch adult population. The sample consists of approximately 2500 respondents age 16 and older, but due to the focus on pensions the *Pensioenbarometer* surveys are only elicited from respondents who are older than 24. All CentERpanel questionnaires are administered via the Internet and members of the panel without Internet access are provided with a set-top box to maintain representativeness.

The 2011 *Pensioenbarometer* was distributed to 2,396 potential respondents and was returned by 1,577 of them (65.8 percent survey response). Item non-response to the subjective survival questions is not an issue: 94.6 percent of the panel members who filled out the questionnaire provided an answer to all survival questions. Furthermore, violations of the arithmetic of probabilities are rare despite the fact that no safeguards were applied to ensure logical consistency of the reported probabilities: 97.1 percent of the complete responses are weakly monotonically decreasing (that is: the probabilities weakly decrease for older age thresholds). Because of these observations, we do not model item non-response or logical inconsistencies. After removing incomplete or logically inconsistent response sequences, we are left with 1,381 observations.

As for demographic characteristics, men and women are evenly represented and the average age is 56. Three quarters of the sample lives with a partner. The sample is mostly healthy: 75 percent rates their own health either as "excellent" or "very good", while only 8 percent is in "bad" health. 41 Percent of the sample has finished some form of higher education (university or an applied college). A further breakdown by sex shows that the sample is better educated than the Dutch average: 44 percent of men and 38 percent of women have completed higher education compared to nationwide averages of 31 and 26 percent respectively in 2009. The average gross personal income in our sample is close to that of the population at large: the economically active within the sample earn 2,978 euro per month compared with a national average of 2,900 euro in 2010.[4]

Table 2 presents descriptive statistics of the reported probabilities and corresponding probabilities from life tables assembled by Statistics Netherlands. Due to the age-eligibility criteria described in section 2.1, the sample sizes are larger for questions referring to older ages (1,157 respondents, 84 percent of the sample, answer all five questions). The average reported probability of survival decreases with the age thresholds from around 75% for age 70 to 27% for age 90. Moreover, the average reported probabilities of men and women are very similar for all age thresholds, whereas the average life table probabilities differ much more between the sexes. Table 2 shows that men on average underestimate their probability of living past ages 70, 75 and 80 by roughly 10 percentage points, while they overestimate their probability of living past age 90 by 5 percentage points. Women, on the other hand, report probabilities that are 4-15 percentage points lower on average than their life table counterparts for all target ages. This suggests that the average life expectancy of men is more in line with that reported in life tables than that of women. Respondents use the full response scale for all questions.

We illustrate the importance of rounding in Table 3. The top panel of Table 3 contains a

[4] A table with descriptive statistics of demographic covariates is available on request.

**Table 2:** Descriptive statistics of the reported survival probabilities and life table (LT) probabilities

| | Men | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | Reported probabilities | | | | | | |
| | N | Mean LT | Mean | S. D. | Min | Q25 | Mdn | Q75 | Max |
| Age 70 | 631 | 87.02 | 76.32 | 21.00 | 0 | 65 | 80 | 90 | 100 |
| Age 75 | 696 | 78.05 | 67.15 | 23.67 | 0 | 50 | 70 | 85 | 100 |
| Age 80 | 742 | 63.14 | 54.49 | 25.75 | 0 | 40 | 50 | 75 | 100 |
| Age 85 | 776 | 43.64 | 40.64 | 26.23 | 0 | 20 | 40 | 60 | 100 |
| Age 90 | 780 | 21.50 | 26.75 | 24.92 | 0 | 5 | 20 | 50 | 100 |
| | Women | | | | | | | | |
| | | | Reported probabilities | | | | | | |
| | N | Mean LT | Mean | S. D. | Min | Q25 | Mdn | Q75 | Max |
| Age 70 | 526 | 90.92 | 75.32 | 21.51 | 0 | 60 | 80 | 90 | 100 |
| Age 75 | 566 | 85.11 | 67.75 | 23.73 | 0 | 50 | 71 | 88 | 100 |
| Age 80 | 593 | 75.11 | 56.18 | 26.20 | 0 | 40 | 60 | 75 | 100 |
| Age 85 | 598 | 58.15 | 42.67 | 26.75 | 0 | 20 | 50 | 60 | 100 |
| Age 90 | 601 | 35.32 | 27.70 | 25.63 | 0 | 5 | 20 | 50 | 100 |

tabulation of the different categories of rounding according to the common rounding scheme described in section 2.5. Frequencies are presented for the entire sample, by sex and by level of education. Rounding appears quite important: about half of the sample reports only probabilities that are multiples of 5, with at least one probability that is not a multiple of 10. Another third of the sample provides probabilities that are consistent with rounding to multiples of 10. Cruder forms of rounding do occur but are rare: 2 percent of the reported sequences consist entirely of probabilities equal to zero or one hundred and another 4 percent consist only zeros, fifties and one hundreds. Focal 50/50s are not likely to be an important concern for our data, since no more than 2.5 percent of the respondents answer 50 percent to all questions. Confirming the analysis in Manski and Molinari (2010), we find some evidence that respondents may round probabilities near the extremes of zero and one hundred differently since 8 percent of the sample reports multiples of 1 near the extremities of the scale. Only 3 percent of the responses are incompatible with all other forms of rounding and are thus interpreted as exact answers (though one might argue that these are rounded to a multiple of 1, the most precise level of rounding allowed for in the questionnaire). We do not find any gender or education pattern in rounding behavior, as can be seen in the remaining columns of the top panel of Table 3.

**Table 3:** Incidence of rounding

| | Common rounding rule (rounding at the level of the individual respondent) | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | By education | | |
| | Frequency | % | Men (%) | Women (%) | Low (%) | Middle (%) | High (%) |
| All 0 or 100 | 27 | 1.96 | 2.05 | 1.83 | 1.92 | 2.53 | 1.58 |
| All 0, 50 or 100 | 52 | 3.77 | 3.72 | 3.83 | 2.64 | 5.57 | 3.34 |
| All multiples of 10 | 448 | 32.44 | 31.54 | 33.61 | 34.86 | 30.38 | 32.16 |
| All multiples of 5 | 710 | 51.41 | 52.31 | 50.25 | 50.96 | 51.39 | 51.85 |
| Some in [1,4] or [96,99] | 105 | 7.60 | 7.69 | 7.49 | 6.49 | 8.10 | 7.91 |
| Other | 39 | 2.82 | 2.69 | 3.00 | 3.13 | 2.03 | 3.16 |
| Total | 1,381 | 100 | 100 | 100 | 100 | 100 | 100 |
| | General rounding rule (rounding at the level of individual probability) | | | | | | |
| | | | By age threshold | | | | |
| Multiples of... | Frequency | % | 70 (%) | 75 (%) | 80 (%) | 85 (%) | 90 (%) |
| ...100 | 717 | 11.02 | 17.20 | 10.30 | 6.89 | 7.79 | 13.69 |
| ...50 | 933 | 14.33 | 14.17 | 11.89 | 16.40 | 16.74 | 12.31 |
| ...25 | 574 | 8.82 | 7.17 | 11.65 | 9.14 | 9.10 | 7.02 |
| ...10 | 3,274 | 50.30 | 49.70 | 49.29 | 54.68 | 49.34 | 48.44 |
| ...5 | 808 | 12.41 | 9.59 | 13.79 | 11.01 | 14.34 | 12.96 |
| ...1 | 203 | 3.12 | 2.16 | 3.09 | 1.87 | 2.69 | 5.58 |
| Total | 6,509 | 100 | 100 | 100 | 100 | 100 | 100 |

The numbers in this table are frequencies of the reported probabilities that fall in the various rounding categories.

The bottom panel of Table 3 describes rounding according to the general rounding scheme explained in section 2.5. The general rounding scheme also indicates substantial rounding, since 50 percent of reported probabilities are multiples of 10 (but not of 50 or 100). Another 14 percent of probabilities are equal to 50 and only 3 percent can only result from rounding to multiples of 1. The variation in frequencies of the different rounding categories across the age thresholds accords with intuition. Respondents express greater certainty, report a higher fraction of zeros and one hundreds, near the extreme ends of the age range. Conversely, 50/50s are more prevalent at the age thresholds of 80 and 85, expressing uncertainty about surviving past those ages.

# 4 Results

## 4.1 Point- and interval identification of life expectancy

Table 4 contains descriptive statistics of the expected age of death calculated from the individual-specific parametric survival curves and the spline survival functions. Men expect to live to the age of 82 on average, which is slightly below the average prediction of 82.85 years found in the life tables. For women we find a larger discrepancy between the average subjective life expectancy and the average actuarial forecast: women expect to live to age 82-83 while the average actuarial prediction is 85.72. This larger discrepancy for women was also observed by Perozek (2008) for the HRS and Kutlu and Kalwij (2012) for the Netherlands. If we divide the sample in age groups, we find that men and women of all ages expect to live shorter than predicted in the actuarial tables. We will analyze the statistical significance of the differences between average subjective life expectancy and the forecasts in life tables in section 4.3 below.

Unsurprisingly, subjective mortality expectations exhibit much more variation than the life tables (the latter only vary with gender and age). The standard deviations of the expected age of death are around 8, while that of the actuarial forecasts is 2.15 for men and 1.42 for women. In the next subsection we analyze whether this variation in expectations is related to covariates such as health and socio-economic status.

The life expectancies calculated from parametric models are very similar to those derived from spline survival functions: all correlations between the expected ages of death are above

**Table 4:** Point estimates of life expectancy

| | Men | | | | | |
|---|---|---|---|---|---|---|
| | | By age bracket | | | | |
| | Overall | 25-35 | 36-45 | 46-55 | 56-65 | 66+ |
| *Parametric approximations* | | | | | | |
| Gompertz | 81.71 | 81.90 | 80.94 | 79.73 | 80.95 | 84.48 |
| | (7.61) | (7.43) | (10.48) | (8.21) | (6.98) | (4.98) |
| Weibull | 82.34 | 83.10 | 82.16 | 80.35 | 81.40 | 84.93 |
| | (7.65) | (7.65) | (10.65) | (8.19) | (7.10) | (4.94) |
| *Non-parametric approximations* | | | | | | |
| Linear splines | 82.11 | 80.72 | 79.08 | 79.59 | 82.02 | 85.94 |
| | (8.09) | (7.78) | (9.29) | (8.10) | (7.91) | (5.99) |
| Cubic splines | 81.95 | 81.24 | 79.33 | 79.40 | 81.82 | 85.55 |
| | (8.29) | (7.76) | (9.62) | (8.36) | (8.30) | (6.12) |
| *Life table forecasts* | | | | | | |
| Statistics Netherlands life tables | 82.85 | 80.56 | 80.84 | 81.39 | 82.61 | 85.62 |
| | (2.15) | (0.07) | (0.11) | (0.23) | (0.47) | (1.91) |
| N | 780 | 36 | 104 | 185 | 233 | 222 |
| | Women | | | | | |
| | | By age bracket | | | | |
| | Overall | 25-35 | 36-45 | 46-55 | 56-65 | 66+ |
| *Parametric approximations* | | | | | | |
| Gompertz | 81.99 | 81.35 | 83.08 | 80.73 | 80.76 | 84.63 |
| | (8.39) | (9.86) | (10.77) | (8.68) | (7.05) | (5.51) |
| Weibull | 82.63 | 82.52 | 83.85 | 81.40 | 81.25 | 85.07 |
| | (8.38) | (9.77) | (10.64) | (8.74) | (7.17) | (5.49) |
| *Non-parametric approximations* | | | | | | |
| Linear splines | 81.78 | 79.29 | 80.72 | 80.28 | 81.61 | 86.03 |
| | (8.33) | (9.18) | (8.91) | (8.27) | (7.86) | (6.51) |
| Cubic splines | 81.61 | 79.32 | 80.73 | 80.10 | 81.33 | 85.76 |
| | (8.61) | (9.78) | (9.22) | (8.65) | (8.04) | (6.80) |
| *Life table forecasts* | | | | | | |
| Statistics Netherlands life tables | 85.72 | 84.29 | 84.49 | 84.95 | 86.04 | 87.99 |
| | (1.42) | (0.04) | (0.08) | (0.21) | (0.37) | (1.24) |
| N | 601 | 61 | 107 | 137 | 178 | 118 |

Reported numbers are averages; standard deviations in parentheses.

0.98. Hence, given sufficiently rich data, computed mortality expectations are robust with respect to the choice for a (non-)parametric model for subjective survival. This finding is in line with De Bresser and Van Soest (2013a), who document that the first two moments of their subjective distributions of the expected replacement rate of income at retirement are very similar for a parametric (log-normal) specification versus cubic splines.

Now we turn to the bounds that we compute using the methodology described in sections 2.4 and 2.5. Table 5 presents sample averages of the bounds on life expectancy. The table reports descriptives for the case in which we do not interpolate expectations (*no smoothing*) and for that in which we do interpolate using cubic splines (*smoothing using cubic splines*). Table 5 contains averages for bounds under the assumptions of no rounding (columns under *no*), and under the common and general rounding rules described in section 2.5.1 (columns under *common* and *general* respectively). Analogous results based on linear spline interpolation are similar to those reported in Table 5 and are available on request.

According to the bounds without smoothing, which do not impose any restrictions on expectations beyond being consistent with the data, both men and women expect to live to age 76-88 on average. Expectations are conditional on the current age of the respondent, so the lower bounds of the intervals for life expectancy increase with age. The upper bound, on the other hand, appears to follow a U-shaped pattern and is the lowest for the 46-55 age bracket. In addition to summary statistics of the bounds, Table 5 also contain information on the average width of the intervals for life expectancy and of the relative decrease in width compared to the uninformative interval (which sets the minimum life expectancy of respondent $i$ to zero and the maximum to $T_{\max} - age_i$, where $T_{\max} = 110$). The former provides an absolute measure of the informativeness of the bounds, while the latter takes into account that the width of the uninformative interval is smaller for older respondents (since the maximum lifespan does not vary with age). The intervals for life expectancy are about 12 years wide on average which translates into a reduction relative to the uninformative interval

**Table 5:** Sample averages of bounds on life expectancy derived in absence of rounding and under common and general rounding

| | Men | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Overall | | | By age bracket | | | | |
| | | | | 25-35 | 36-45 | 46-55 | 56-65 | 66+ |
| | No | Common | General | No | No | No | No | No |
| *No smoothing* | | | | | | | | |
| LB | 76.43 | 75.03 | 73.45 | 72.46 | 70.93 | 73.31 | 77.12 | 81.55 |
| UB | 87.91 | 89.41 | 91.32 | 88.98 | 87.22 | 85.87 | 86.91 | 90.45 |
| Width | 11.37 | 14.39 | 17.87 | 16.52 | 16.29 | 12.56 | 9.80 | 8.90 |
| Percentage decrease[a] | 77.88 | 72.13 | 65.31 | 78.80 | 76.42 | 78.90 | 80.13 | 75.22 |
| N | | 780 | | 36 | 104 | 185 | 233 | 222 |
| *Smoothing using cubic splines* | | | | | | | | |
| LB | – | 80.44 | 78.19 | – | – | – | – | – |
| UB | – | 83.48 | 85.34 | – | – | – | – | – |
| Width | – | 3.04 | 7.14 | – | – | – | – | – |
| Percentage decrease[a] | – | 94.18 | 86.35 | – | – | – | – | – |
| N | | 778 | 611 | | | | | |
| | Women | | | | | | | |
| | Overall | | | By age bracket | | | | |
| | | | | 25-35 | 36-45 | 46-55 | 56-65 | 66+ |
| | No | Common | General | No | No | No | No | No |
| *No smoothing* | | | | | | | | |
| LB | 75.66 | 73.99 | 72.20 | 70.08 | 72.66 | 73.91 | 76.83 | 81.52 |
| UB | 87.93 | 89.59 | 91.68 | 88.49 | 88.77 | 86.66 | 86.40 | 90.65 |
| Width | 12.27 | 15.60 | 19.48 | 18.41 | 16.12 | 12.74 | 9.57 | 9.13 |
| Percentage decrease[a] | 78.09 | 72.27 | 65.15 | 76.42 | 76.87 | 78.75 | 80.59 | 75.51 |
| N | | 601 | | 61 | 107 | 137 | 178 | 118 |
| *Smoothing using cubic splines* | | | | | | | | |
| LB | – | 79.90 | 77.44 | – | – | – | – | – |
| UB | – | 83.29 | 85.51 | – | – | – | – | – |
| Width | – | 3.38 | 8.07 | – | – | – | – | – |
| Percentage decrease[a] | – | 94.10 | 85.65 | – | – | – | – | – |
| N | | 600 | 463 | | | | | |

[a] Percentage decrease in width of the interval for life expectancy relative to the uninformative interval. The uninformative interval of respondent $i$ is equal to $110 - age_i$.

of 78 percent. The average absolute width drops with age from 17 years for the youngest age group to 9 years for those of age 66 and older. The percentage decrease in width relative to the uninformative interval does not show any clear pattern with age.

By definition allowing for rounding makes the estimated intervals wider and thus less informative. Assuming a common rounding rule results in bounds with an average width of 15 years, or an average reduction from the uninformative interval of 72 percent. The general rounding scheme yields slightly wider intervals with an average width of 18 years.

We obtain much tighter bounds on life expectancy if we interpolate expectations between the elicited survival probabilities. Interpolation allows one to focus on the ambiguity that results from rounding, isolating it from the coarseness inherent in eliciting expectations by means of a limited number of probabilities. Under the common rounding assumption, interpolation reduces the average width of the interval from 15 years to 3 years. Though not reported in Table 5, the reduction is apparent in all age groups. The average for the 25-36 year olds, for instance, is diminished from 22 to 5 years, while that among respondents aged 66 and older drops from 11 to 2 years.

## 4.2 Life expectancy and demographic variables

Next we investigate the relationship between life expectancy and demographic variables. We estimate linear regressions with the point identified life expectancies as dependent variables. For the bounds we apply partially identified models according to the methods presented in Imbens and Manski (2004).[5] Estimation results are presented in Table 6. The model specifications in that table pool men and women, because we could not reject the null hypothesis of equal coefficients for the sexes. Point identified models show that *age* and self-reported *health* are the most important covariates of subjective life expectancy. As expected, remaining life expectancy decreases non-linearly with age. Health too is strongly related to

[5]We estimate the interval regressions using the Stata program CI1D, presented in Beresteanu et al. (2010).

subjective life expectancy: compared to the baseline of people in excellent health, those in bad health expect to live 7 years shorter on average. Respondents in fair or good health also expect to live shorter than their healthier peers. Like Kutlu and Kalwij (2012), we do not find significant associations between life expectancy and education and income if we also condition on subjective health. This lack of an association between socio-economic status and subjective mortality, conditional on subjective health, is plausible in the context of the Dutch healthcare system, because the quality of medical services is the roughly same for everybody. However, poorly educated and income-poor individuals are especially likely to be in worse health, partly because they are more likely to engage in behaviors that affect their health negatively (like smoking and drinking). Therefore, we find that respondents from the lowest income group and those with the poorest education expect to die younger if we remove subjective health from the estimated equation (estimates available on request). Note that the estimates are not sensitive to the way in which expectations are approximated: similar conclusions emerge whether we fit Weibull distributions or cubic splines.

The rightmost columns of Table 6 present estimates from models with interval-censored life expectancy as the dependent variable. The estimates from partially identified models estimated on the bounds without smoothing show that little can be learned about differences in life expectancy across the sample if one is unwilling to interpolate expectations between data points. There are clear differences between the identified sets that are in line with the coefficient estimates from point identified models. For instance, the set-estimate of the coefficient of being in bad health, which is associated with a 7-year lower life expectancy relative to the baseline in point identified models, is $(-23.7; 8.9)$. Though that interval suggests a negative correlation, zero is included in all identified sets (and by implication in all 95% confidence sets) except for the constant. Hence, we are not able to draw conclusions about the sign of any of the coefficients: a clear indication that the bounds are too wide for useful inference. Since rounding only makes the bounds on life expectancy wider, this holds

**Table 6:** Point and partially identified models of remaining life expectancy

| | | | Partially identified models | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | No smoothing | | Cubic splines | | | |
| | Point identified models | | No rounding | | Common rounding | | General rounding | |
| | Weibull | Cubic spline | LB | UB | LB | UB | LB | UB |
| Age | -1.740*** | -1.539*** | -3.278 | 2.256 | -2.257 | -0.796 | -3.176 | 0.266 |
| | (0.106) | (0.114) | (-3.526; 2.504) | | (-2.475; -0.578)** | | (-3.446; 0.535) | |
| Age squared/100 | 0.761*** | 0.670*** | -1.762 | 3.079 | 0.0264 | 1.290 | -0.883 | 2.103 |
| | (0.0890) | (0.0978) | (-1.981; 3.298) | | (-0.144; 1.461) | | (-1.115; 2.334) | |
| Male | -0.808* | -0.616 | -12.571 | 11.185 | -3.802 | 2.663 | -8.397 | 6.935 |
| | (0.430) | (0.441) | (-13.423; 12.037) | | (-4.687; 3.547) | | (-9.436; 7.974) | |
| Educ. - primary school | -1.092 | -0.789 | -13.151 | 11.713 | -3.811 | 1.940 | -9.338 | 5.529 |
| | (0.937) | (1.020) | (-14.734; 13.297) | | (-5.527; 3.655) | | (-11.642; 7.833) | |
| Educ. - higher secondary | -0.239 | -0.0520 | -12.898 | 12.681 | -3.602 | 3.062 | -8.626 | 7.707 |
| | (0.735) | (0.781) | (-14.184; 13.968) | | (-5.210; 4.671) | | (-10.344; 9.425) | |
| Educ. - lower vocational | 0.143 | -0.144 | -14.136 | 13.911 | -4.349 | 3.811 | -9.803 | 9.065 |
| | (0.696) | (0.705) | (-15.540; 15.314) | | (-5.717; 5.179) | | (-11.482; 10.744) | |
| Educ. - higher vocational | -0.515 | -0.415 | -13.706 | 12.771 | -4.008 | 2.939 | -9.247 | 7.592 |
| | (0.575) | (0.592) | (-14.861; 13.926) | | (-5.134; 4.064) | | (-10.643; 8.988) | |
| Educ. - university | 0.456 | 0.164 | -14.281 | 14.712 | -3.961 | 3.808 | -9.151 | 9.085 |
| | (0.680) | (0.674) | (-15.609; 16.040) | | (-5.192; 5.039) | | (-10.822; 10.755) | |
| Income < 1150 euro/month | -0.336 | -0.438 | -15.306 | 14.327 | -4.586 | 3.708 | -9.628 | 9.741 |
| | (1.125) | (1.087) | (-17.247; 16.268) | | (-6.639; 5.762) | | (-12.091; 12.203) | |
| Income 1801-2600 euro/month | -0.282 | -0.120 | -12.737 | 12.744 | -3.724 | 3.291 | -8.101 | 8.252 |
| | (0.703) | (0.716) | (-13.908; 13.915) | | (-4.928; 4.495) | | (-9.816; 9.968) | |
| Income > 2600 euro/month | -0.331 | -0.0576 | -13.122 | 13.059 | -3.637 | 3.343 | -8.263 | 8.421 |
| | (0.668) | (0.671) | (-14.461; 14.398) | | (-4.890; 4.596) | | (-9.965; 10.122) | |
| Health - bad | -6.713*** | -7.668*** | -23.717 | 8.920 | -11.606 | -2.449 | -16.660 | 3.988 |
| | (1.039) | (1.082) | (-25.682; 10.886) | | (-13.736; -0.319)** | | (-19.009; 6.337) | |
| Health - fair | -5.378*** | -6.351*** | -21.302 | 8.746 | -10.190 | -2.236 | -14.821 | 3.312 |
| | (0.764) | (0.798) | (-22.797; 10.242) | | (-11.764; -0.662)** | | (-16.661; 5.152) | |
| Health - good | -1.366** | -2.155*** | -15.403 | 11.416 | -5.773 | 1.670 | -10.151 | 6.549 |
| | (0.660) | (0.697) | (-16.734; 12.746) | | (-7.092; 2.989) | | (-11.820; 8.217) | |
| Constant | 101.9*** | 93.39*** | 14.901 | 169.661 | 72.721 | 113.506 | 41.289 | 139.392 |
| | (3.249) | (3.396) | (7.710; 176.852)** | | (66.029; 120.198)** | | (32.937; 147.744)** | |
| N | 1,380 | 1,380 | 1,380 | | 1.377 | | 1,073 | |
| R-squared | 0.735 | 0.680 | – | | – | | – | |

Standard errors in parentheses for point identified models
95% confidence sets in parentheses for the partially identified models
***significant at 1%; **significant at 5%; *significant at 10%

even more strongly if we allow for rounding. Without additional assumptions there is not enough information in the data to do precise inference on differences in life expectancy across different socio-demographic groups.

Table 6 shows that if we do smooth expectations by means of cubic splines and allow for rounding according to the common rounding rule, we find evidence for similar correlations that are apparent in point identified models. In particular, zero is not included in the 95% confidence sets for the coefficients on the indicators for bad and fair health. These results show that the limited number of elicited points on the survival functions is a more important limitation on the informativeness of the data than (common) rounding. However, if we simultaneously interpolate and allow for general, worst-case, rounding, the identified sets again become too wide to draw any conclusions (though they remain narrower than in the no-smoothing, no-rounding case). Therefore, the informativeness of the data hinges not only on our willingness to smooth beliefs, but also on the particular type of rounding that we assume. Finally, note that the sample sizes for the partially identified models of smoothed expectations are smaller than those for the other models. This is due to the fact that the cubic spline functions that trace the upper and lower bounds for the true, non-rounded probabilities sometimes cross each other, in which case we drop that observation. This is rare for the case of common rounding, only 3 observations are lost this way, but quite common for general rounding for which we lose 307 observations (or 22% of the sample). This complication does not occur with linear splines, so we verified the patterns from Table 6 using linear splines. All estimates are similar to those reported here, results are available on request.

## 4.3 Consistency of expectations with life tables

As mentioned in the introduction, consistency of subjective expectations with published life-tables has been an active area of research since the first subjective longevity questions were posed in the HRS in 1992 (Gan et al., 2005). Most recently, Perozek (2008) found that in the 1992 wave of the HRS women expected to live shorter than the 1992 actuarial forecasts predicted and that these expectations were in line with subsequent downward revisions in forecasts of female life expectancy (though even the most recent forecasts still implied longer life expectancy compared to the expectations). We now investigate the correspondence between the Dutch life tables assembled by Statistics Netherlands, as of December 2010, and subjective expectations. First we look at the point identified life expectancies. Figure 6 presents kernel regressions of subjective life expectancy, computed from cubic splines, and the actuarial forecast. Analogous figures based on Weibull or Gompertz distributions and linear splines are similar and are available upon request. The left panel of Figure 6 shows that for men expectations and actuarial forecasts are very much in line on average: the actuarial tables lie within the 95% confidence interval for the average subjective life expectancy for (almost) all ages. For women, on the other hand, we find that official forecasts are above the 95% confidence interval for all ages between 30 and 70 (see the right graph in Figure 6). Hence, women indeed expect to die significantly younger than the actuarial estimates predict. The size of this difference is large: close to 5 years around the age of 60. Note, however, that these estimates do not take into account missing information between elicited probabilities or rounding.

The top panels of Figure 7 plots 95% confidence bands for subjective remaining life expectancy for men (left graph) and women (right graph). These bands are based on bounds derived without smoothing expectations and span the width from the lower end of the 95% confidence interval for the lower bound on life expectancy to the higher end of the 95% confidence interval for the upper bound. The solid lines are the corresponding predictions from life-tables. Even without rounding, the data for men are consistent with the actuarial forecasts across all ages. Moreover, allowing for rounding does not affect the average bounds much, though the effect is slightly larger at younger ages. The top right graph of Figure 7 contains the same information for women. Contrary to the findings in Figure 6, we cannot

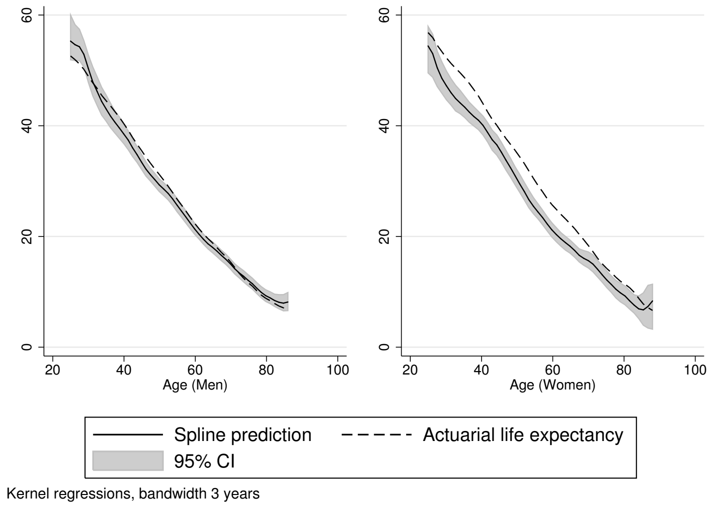


**Figure 6:** Actuarial forecasts and subjective life expectancy (expectations approximated using cubic splines)

reject that beliefs of women are on average consistent with the life tables once we take into account that we only observe a few points on the subjective survival function. Though we find that the actuarial forecasts from Statistics Netherlands are on the high end of the confidence interval for subjective life expectancy, they are between the bounds that do not allow for rounding. Even around age 60, where Figure 6 indicates that the difference between life tables and expectations is substantial, the non-parametric bounds do not allow us to conclude that the actuarial forecasts are above the average upper bound on subjective life expectancy. When we take rounding into account, the actuarial figures lie well within the interval for all ages. Hence, we conclude that without additional assumptions we cannot reject the hypothesis that expectations are on average consistent with the life tables for both men and women in our sample. Robustness checks indicate that these conclusions remain largely unchanged when we lower the maximum lifespan to 100 years: for men the actuarial forecasts remain well within the subjective bounds for all ages, while for women the actuarial forecasts remain within the 95% confidence bands if we allow for rounding.[6]

[6]Estimates available upon request.

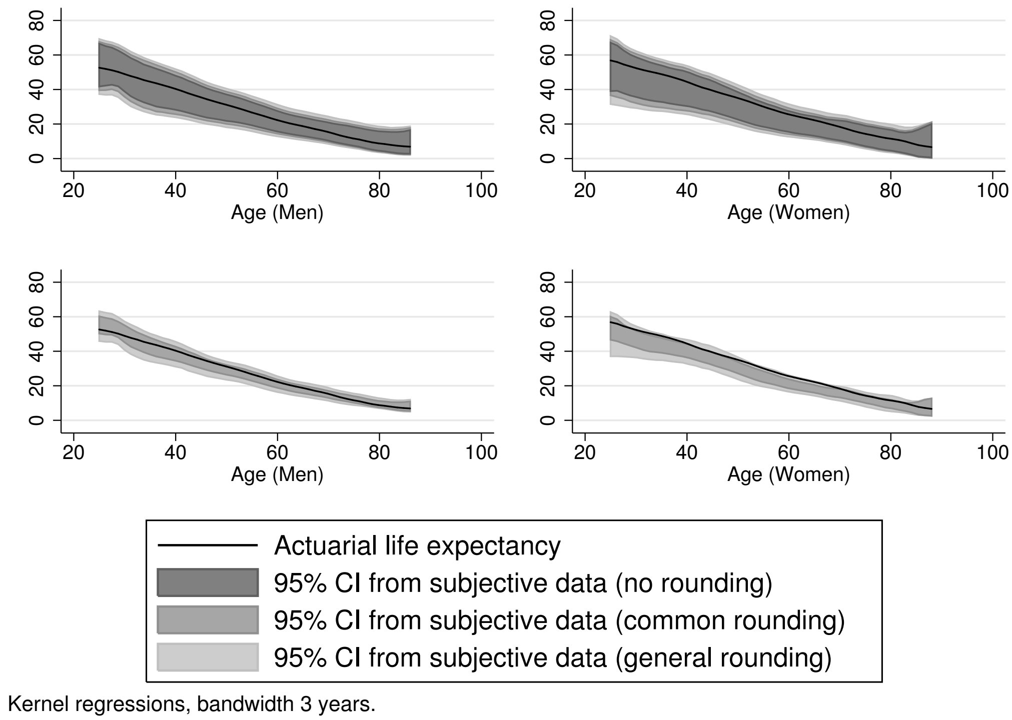


**Figure 7:** Non-parametric bounds on life expectancy with (top panels) and without (bottom panels) interpolation between reported probabilities

In the previous sections we have shown that the width of the bounds can be reduced substantially by simultaneously allowing for rounding of the reported probabilities and interpolating expectations between the elicited points on the subjective survival functions. A natural question is whether rounding alone can close the gap between average expectations and the actuarial forecasts for women that is evident in Figure 6. The bottom right graph in Figure 7 shows that common rounding closes the gap for all ages except for 50 to 70 year olds and around the age of 30, for which small differences remain. The more conservative general rounding scheme, on the other hand, removes those discrepancies. Under general rounding, we cannot reject the null hypothesis that the average upper bound of the intervals is equal to the corresponding life table forecast for any age. The same conclusions emerge if we approximate expectations by means of linear splines.

# 5 Conclusion

When investigating the subjective expectations held by survey respondents, researchers usually point identify beliefs by means of parametric specifications for expectations or non-parametric interpolation between data points (e.g. Dominitz and Manski, 1997; Dominitz, 1998; Perozek, 2008; Kutlu and Kalwij, 2012; Bellemare et al., 2012). Building on the work of Manski (2003); Engelberg et al. (2009); and Manski and Molinari (2010), we analyze what can be learned about subjective life expectancy under weaker assumptions. In particular, we propose two methods to partially identify expectations, constructing identified sets within which subjective survival functions must fall to be consistent with the data. The first is a "worst-case" scenario in which we neither assume a functional form for expectations nor do we interpolate between the elicited points on the subjective survival functions. We show that this procedure can easily be generalized to allow for rounding of the subjective probabilities reported in surveys. The second approach focuses exclusively on the ambiguity in the data that is due to rounding of the reported probabilities. This method interpolates expectations between the elicited points, which in combination with an assumed rounding scheme yields identified regions for expectations that are much narrower than those that do not interpolate. Both methods we propose construct bounds on relevant aspects of beliefs.

Our application concerns mortality expectations from a representative sample of Dutch adults. We find that subjective life expectancies are similar regardless of whether we approximate expectations by parametric models or by (linear or cubic) spline interpolation: correlations between the resulting life expectancies are above 0.98.

If we do not interpolate expectations between elicited points on the survival functions, the bounds on life expectancy are too wide to be informative. In particular, models for interval-censored dependent variables fail to corroborate the associations between life expectancy and the covariates *age* and *health* that are highly statistically significant in non-censored models. Rounding makes the intervals even less informative. However, if we do smooth survival functions between observed points, the intervals can be narrowed enough to allow for useful inference. Under the assumption that all probabilities reported by a given respondent are rounded to the same extent, the same assumption made by Manski and Molinari (2010), partially identified models show the same patterns that emerge from point identified models. Allowing for more conservative rounding, every probability is rounded to the maximum extent, again yields wide and uninformative intervals.

We match our subjective life expectancies to life tables constructed by Statistics Netherlands. If we point identify expectations, either parametrically or non-parametrically, we confirm for the Netherlands the finding from Perozek (2008) that women in the US expect to live shorter on average than the published life-tables predict. The difference is largest around the age of 60. Men, on the other hand, have expectations that are on average in line with the actuarial forecasts. The bounds on life expectancy do not allow us to reject that beliefs held by men are consistent with life-tables in the aggregate, regardless of whether or not we interpolate between data points. In contrast to the case in which we point identify expectations, we can no longer reject that expectations of women are consistent with the forecasts if we do not smooth expectations. This emerges even starker if we allow for rounding in the reported probabilities. If we interpolate expectations and simultaneously allow for common rounding, expectations of women are inconsistent with the life tables around the ages of 30 and 60 (the size of the remaining difference is much smaller). If we allow for the more conservative rounding scheme that does not impose common rounding, we cannot reject consistency of women's expectations with actuarial forecasts for any age.

The general idea that emerges from this paper is that it is possible to learn about subjective expectations without imposing parametric restrictions on beliefs or even point identifying them. The methods we propose are sufficiently flexible to take into account rounding issues that are relevant for survey data of many types. Moreover, they can be combined with plausible assumptions on beliefs to allow more precise inference that remains less restrictive than fully characterizing expectations by parametric distributions. Our partial identification framework yields new insights into the influence of parametric assumptions on the analysis of this important and increasingly popular type of data.

# References

Bellemare, C., Bissonnette, L., and Kröger, S. (2012). Flexible approximation of subjective expectations using probability questions: an application to the investment game. *Journal of Business and Economic Statistics*, 30(1):125–131.

Bellemare, C., Kröger, S., and Van Soest, A. (2008). Measuring inequity aversion in a heterogeneous population using experimental decisions and subjective probabilities. *Econometrica*, 76(4):815–839.

Beresteanu, A. and Molinari, F. (2008). Asymptotic properties for a class of partially identified models. *Econometrica*, 76(4):763–814.

Beresteanu, A., Molinari, F., and Steeg Morris, D. (2010). Asymptotics for partially identified models in STATA. Stata code for BLP on interval data.

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. (2013a). Retirement expectations and satisfaction with retirement provisions. Forthcoming in *Review of Income and Wealth*.

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

Dominitz, J. (1998). Earnings expectations, revisions and realizations. *Review of Economics and Statistics*, 80(3):374–388.

Dominitz, J. and Manski, C. F. (1997). Using expectations data to study subjective income expectations. *Journal of the American Statistical Association*, 92(439):855–867.

Dominitz, J. and Manski, C. F. (2006). Measuring pension benefit expectations probabilistically. *Labour*, 20(2):201–236.

Engelberg, J., Manski, C. F., and Williams, J. (2009). Comparing the point predictions and subjective probability distributions of professional forecasters. *Journal of Economic and Business Statistics*, 27(1):30–41.

Gan, L., Hurd, M. D., and McFadden, D. L. (2005). Individual subjective survival curves. In Wise, D. A., editor, *Analyses in the economics of aging*. University of Chicago Press, Chicago.

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.

Hurd, M., Smith, J. P., and Zissimopoulos, J. M. (2004). The effects of subjective survival on retirement and social security claiming. *Journal of Applied Econometrics*, 19(6):761–775.

Imbens, G. and Manski, C. F. (2004). Confidence intervals for partially identified parameters. *Econometrica*, 72(6):1845–1857.

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

Manski, C. F. (2003). *Partial Identification of Probability Distributions*. Springer, New York.

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.

Peracchi, F. and Perotti, V. (2011). Subjective survival probabilities and life tables: evidence from europe. EIEF working paper 16/10.

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.

Tamer, E. (2010). Partial identification in econometrics. *Annual Review of Economics*, 2(1):167–195.

Van der Klaauw, W. (2012). On the use of expectations data in estimating structural dynamic choice models. *Journal of Labor Economics*, 30(3):521–554.

Van der Klaauw, W. and Wolpin, K. (2008). Social security and the retirement and savings behavior of low-income households. *Journal of Econometrics*, 145(1-2):21–42.