Estimating Knightian Uncertainty from Survival Probability Questions on the HRS*

Daniel Hill, Michael Perry and Robert J. Willis
University of Michigan

Abstract

Since Frank Knight (1921) introduced the distinction, economists have recognized that risk is the special case of uncertainty in which probabilities are known. Based on subjective survival probability questions in the Health and Retirement Study we use an econometric model to estimate the determinants of individual-level uncertainty about personal longevity. This model is built around the Modal Response Hypothesis—a mathematical expression of the idea that survey responses of zero, one-half or one to probability questions indicate a high level of uncertainty about the relevant probability. Our results indicate that poor cognition and poor health are associated with increases in uncertainty about survival.

To be Presented at
"Labor and Social Interaction in Economics"
A Conference in Honor of Yoram Weiss' Sixty-Fifth Birthday
The Eitan Berglas School of Economics, Tel-Aviv University
December 16 th -17 th, 2004

---

*A highly preliminary and incomplete versions of this paper was presented in a meeting organized by Daniel McFadden for the 6th Invitational Choice Symposium, Estes Park, Colorado, June 5-8, 2004 and at a workshop at the University of Washington, Seattle, on November 19, 2004. This paper is also scheduled to be presented at a session on “Subjective Probability Distributions,”American Economic Association Meetings, Philadelphia, January 6-8, 2005. We are grateful to grant U01-AG09740 from the Behavioral and Social Research Program of the National Institute on Aging for supporting the collection of the HRS data used in this paper and to grant R01-AG020638 for partial support of the research for this paper.

Since Frank Knight (1921) introduced the distinction, economists have recognized that risk is the special case of uncertainty in which probabilities are known. Conventional economic theory, rooted in subjective expected utility (*SEU*) theory, is often interpreted as having erased this distinction so that evidence that individuals dislike uncertainty, such as the Ellsberg Paradox (Ellsberg 1961), is taken to be evidence against SEU theory itself and has provided a rationale for the development of non-expected utility theories. (See Starmer, 2000, for a survey.) However, recently Lillard and Willis (2001) and Kezdi and Willis (2003) have shown that Knightian uncertainty, as distinct from risk, may play an important role within conventional *SEU* theory whenever uncertainty pertains to an environment in which an individual will be subject to repeated risks. For example, in such environments, a mean-preserving spread in subjective uncertainty about the rate of return leads risk averse households to choose portfolios with smaller risk. More generally, increased uncertainty will influence choices whenever probabilities enter the objective function non-linearly.

The goal of this paper is to use data on survival expectations elicited by questions in the Health and Retirement Study (HRS) to estimate the distribution of an individual's beliefs about the probability of survival to a given age.[1] For example, for persons younger than age 70, the survival question that we analyze in this paper is: "Using a number from 0 to 100, what do you think are the chances that you will live to be at least 80?"[2] In our analysis, we will develop and estimate an econometric model which allows

[1] See Manski (2004) for a survey of methods of eliciting subjective probabilities.
[2] For persons older than 69, the target age is increased to 85, 90, 95 and 100 for individuals in successive five year age intervals so that the target is never closer than ten years in the future.

us to use responses to these questions to study the determinants of subjective expected survival probabilities and the degree of subjective uncertainty about these probabilities.

Two major findings from previous studies using the HRS survival probability questions stimulate the approach taken in this paper. First, on average, responses to these questions appear to be remarkably accurate. Average survival probabilities are close to those presented in life tables and co-vary with variables such as smoking, drinking, health conditions and education in ways that would be that would be expected from actual mortality (Hurd and McGarry, 1995). Using panel data from the HRS, several studies have found that respondents modify their probabilities in response to new information such as the onset of a new illness and that the probabilities are useful predictors of actual mortality (Hurd and McGarry, 2002; Gan, Hurd and McFadden, 2003; Smith, Taylor and Sloan, 2001; Siegel, Bradley and Kasl, 2003; Smith et al., 2001) Second, however, at the individual level survival expectations appear to be very noisy with extensive heaping at values of “0”, “50” and “100” (we refer to these responses as “focal”; we refer to non-focal answers as “exact”). Moreover, these two features—reasonable accuracy of probability answers, on average, and extensive heaping on focal values at the individual level—are descriptive of findings for responses to HRS probability questions on a wide variety of topics (Lillard and Willis, 2001).

We seek in this paper to understand how an individual’s response to the survey question is related to his subjective probability beliefs about survival to the target age. We assume that individuals have in mind a subjective distribution over possible values of the probability in question. In some cases, the belief may be precise and could be well represented by a single number. More often, a person’s beliefs may be less precise. We assume these beliefs can be represented by a subjective distribution, *G(p)*, with density, g*(p)*. The expected value of this distribution, $\overline{p} = \int pg(p)dp$, is the probability of the event that would be reported by a Bayesian statistician or which the enters an expected utility function in conventional *SEU* models of decision-making under uncertainty.[3] The spread of g*(p)* measures the degree of imprecision of beliefs or, in Knightian terms, the person's degree of uncertainty which, under certain conditions, may influences decisions within the *SEU* framework.[4]

It would be convenient if answers to HRS survival probability questions could be interpreted as equal to $\overline{p}$. However, as Lillard and Willis (2001) argue, this interpretation is undermined by the fact that HRS probability questions on a wide variety of topics generate a large number of focal answers. If individuals are assumed to have very precise beliefs which can be summarized by a single integer it is difficult to account for persistent heaping on three integers—0, 50 and 100. The same objection holds if they have less precise beliefs and are assumed to average all possible values of the probability,

---

[3] For example, consider a game in which a ball is drawn from an urn containing red and black balls. Payoffs are 100 for a red ball and 0 for a black ball which occur with probabilities *p* and *1*-p, respectively. Since probabilities enter the utility function linearly, expected utility, is independent of the spread of $g(p)$ and depends only on $\overline{p}$. Specifically, $EU = \int [pU(100) + (1-p)U(0)]g(p)dp = \overline{p}U(100) + (1-\overline{p})U(0)$.

[4] Continuing with the example from footnote 3, Lillard and Willis (2001) show that uncertainty about *p* will influence expected utility if the urn game involves repeated draws because utilities are then non-linearly related to *p*. For example, payoffs from two draws from the urn are 200 with probability p^2, 100 with probability $2p(1-p)$ and 0 with probability $(1-p)^2$. It is easy to show that a mean-preserving spread of $g(p)$ causes the distribution of payoffs to become riskier. Hence, a risk-averse person would prefer an urn with less uncertainty. Conversely, if a person could change his choice of urn after observing a drawn, he may prefer a less certain urn because uncertainty increases its option value.

weighted by g*(p)*, before answering the survey question. Moreover, it should be noted that weighted averages are cognitively difficult to compute and that HRS respondents take, on average, only about 15 seconds to answer a given probability question. Presumably, the survey response entails some form of quick "gut answer" of the sort that people give many times a day to equally hard questions on a myriad of topics.

Lillard and Willis (2001) suggest that a less cognitively difficult response would be for an individual to report that probability which is the most likely among all possible probabilities. This is the mode of *g(p)*. Lillard and Willis call this the *modal response hypothesis* (*MRH*). When the degree of uncertainty is relatively low, the mode provides a very good estimate of the mean. Thus, reports of the mode might be construed as a "fast and frugal" algorithm of the sort that the psychologists, Gigerenzer et al. (1999), suggest often help people make complex judgments in situations for which a rapid response is required. As uncertainty grows, however, the mode of $g(p)$ tends to move away from $\overline{p}$, toward zero if $\overline{p} < 1/2$ and toward one if $\overline{p} > 1/2$. When uncertainty becomes sufficiently great, $g(p)$ becomes J-shaped, monotonically increasing in p if $\overline{p} > 1/2$ and monotonically decreasing in p if $\overline{p} < 1/2$. According to the *MRH*, this causes a respondent to report focal answers at "0" or "100" which are, respectively, significantly lower or higher than $\overline{p}$. As uncertainty continues to grow, $g(p)$ acquires a bi-modal U-shape with maxima at zero and one. In this case, the *MRH* says that respondents are sufficiently uncertain that they report "50." In this aspect, the interpretation of "50" answers according the *MRH* is similar to Bruine de Bruine et al. (2000), and Fischhoff and Bruine de Bruin (1999) who suggest that some respondents may report 50 as an expression of *epistemic uncertainty* which is equivalent to the phrase "God only knows".

The major methodological contribution of this paper is to derive a formal econometric model based on the *MRH* and to estimate this model using data on subjective survival probabilities from the HRS. This paper is part of a larger project that aims to explore and test alternative theories of survey response to questions about subjective probability beliefs; to design improved questions to elicit probabilistic expecations; to understand the determinants of probability beliefs in a variety of domains; and to address a range of specific topics on how probability beliefs affect decisions about health, savings and portfolio choice, and retirement. For the purposes of this paper, however, we shall treat the *MRH* as a maintained hypothesis and confine ourselves to the investigation of the determinants of subjective beliefs about survival. Under the *MRH* and additional assumptions to be discussed below, we use cross-sectional data from the HRS survival probability questions to estimate the distribution of beliefs, $G(p \mid Z_p, Z_a)$, where Z_p is a set of variables that affect the median survival probability and Z_a is a set of variables that affect the imprecision or ambiguity of these beliefs.

The paper proceeds as follows. In Section 1, we describe the patterns of focal and non-focal answers to the HRS survival probability question that we analyze in this paper. In Section 2, we discuss the modal response hypothesis and show how it might account for focal responses to subjective probability questions. In Section 3, we develop an econometric model based on the *MRH* that can be estimated by maximum likelihood. This model allows for individual variation in the level of survival risk, interpreted as a person's expected survival probability, and for individual variation in subjective uncertainty about that risk. Estimates of several specifications of the model are presented in Section 4. A discussion of implications for further research is presented in the concluding section.

Estimates of the econometric model indicate that HRS respondents have considerable individual-specific variation in survival risks and in uncertainty about the precise value of this risk. We first estimate a basic version of the model using data from the 2002 wave of HRS in which survival risk is a function of age, sex, race and education and uncertainty depends on these same demographic variables and also on two cognitive measures. The demographic variables affect survival risk in the expected way. Individuals with lower cognitive scores exhibit more uncertainty and uncertainty is sharply greater for older and less educated persons. In previous research, Kezdi and Willis (2003) and Bassett and Lumsdaine (1999) found evidence of a strong "optimism" component in answers to HRS probability questions on a variety of topics, indicating that people who think one nice thing will happen also think that other nice things will happen. Lillard and Willis (2001) also show that, to varying degrees, individuals have a tendency to give focal answers to probability questions across different domains. Lillard and Willis interpret this as an indicator of imprecise probability beliefs. In a second version of the model, we add an optimism component extracted from a factor analysis of HRS probability questions on topics other than survival as a determinant of risk and a measure of the propensity to give focal answers to these questions as a determinant of uncertainty. These measures are both highly significant, indicating that there is a great deal of population heterogeneity in subjective survival beliefs. Finally, in a third version of the model, we investigate how variations in health affect subjective risk and uncertainty, using measures of self-reported health in the 2000 and 2002 waves. As expected, these measures have very large effects on survival risk. We also find that poor health is associated with significant increases in uncertainty. In the conclusion, we speculate about the interpretation of this result and its potential implications for decisions that depend on survival expectations.

1. Focal Answers to Subjective Probability Questions.

Although the subjective probability responses in HRS seem reasonable when averaged across respondents, individual responses appear to contain considerable noise and are often heaped on values of “0”, “50” and “100”. Considering the whole group of probability questions in HRS-1998, for example, while only 5% of respondents refused to answer the probability questions, 52% of questions were heaped on either “0” or “100” and an additional 15% were heaped on “50”.

These patterns are illustrated in Figure 1 by histograms of responses to the HRS-2002 question asking, “On a scale of 0 to 100, what do you think are the chances that you will live to be at least X” where X is a target age at least 10 years in the future. Histograms for persons aged 50-64, 65-74, and 75-90, respectively, are shown in the three panels of Figure 1. Each histogram shows a high frequency of focal answers, especially at 50. However, the average value of subjective probabilities across persons decreases sharply with age, from 66 percent for age 50-64, to 57 percent for age 65-74, to 36 percent for 75-90. Moreover, for the first two age groups, these averages are quite close to the average values of the corresponding probabilities of surviving from an individual’s current age to age X which are 59 percent and 58 percent, respectively, based on the 2001 U.S. Life Table. The 36 percent subjective survival probability reported by the oldest group of 75-90 year olds appears to be quite optimistic relative to a life table estimate of 23 percent.

Some psychologists, especially Fischhoff, Bruine de Bruin and their colleagues (Fischhoff and Bruine de Bruin, 1999; Bruine de Bruin et al., 2000) have argued that answers of "50" may reflect "epistemic uncertainty;" that is, a failure to have any probability belief at all about the event in question or, at least, to have no clear idea of what the probability could be. Alternatively, of course, an answer of "50" might reflect a very precise belief about the probability that a fair coin will come up heads or perhaps a somewhat less precise belief that a given event is about equally likely to occur or not occur. Indeed, while HRS probability questions offer participants a scale of integers from 0 to 100, the large majority of "non-focal" answers are integers ending in "5" or "0", suggesting that responses from most people involve rounding or approximation.[5]

There has been much less emphasis in the psychological literature on focal answers at "0" or "100." When a probability question concerns an event such as the chance of being alive ten or fifteen years from now, it does not seem credible to assume that a respondent who gives such an answer of "100" is completely certain he will be alive then and, apart from a person diagnosed with a terminal illness, an answer of "0" should not be taken at face value, either.

It is possible to regard answers of "0" or "100" as approximations which are no different in kind than rounded answers of "5", "40" or "95". However, in a discussion of Gan, Hurd and McFadden (2003), Willis (forthcoming) provides evidence against this

[5] The current 100 point scale in HRS was introduced in hopes of reducing extensive heaping on focal values of "0" "5" and "10" that were observed with the10 point scale used in HRS-1992. However, the frequency of focal answers actually increased with the 100 point scale and, as just noted, few respondents utilize integers not divisible by 5 (by our calculations).

interpretation in a simple regression of actual mortality by 1995 on individual answers to the subjective survival probability question in 1993 using a sample of persons over 70 from the AHEAD cohort of HRS. He finds that a person who responds “0” has an actual survival probability that is 13 percentage points higher than a person who gives very low non-focal answer. Similarly, a person who responds “100” has a survival chance that is 3.8 percentage points lower than a person who gives a non-focal answer near 100. This suggests that focal answers at the extreme may, like answers at “50”, reflect more imprecise, ambiguous or uncertain probability beliefs than those of persons who give non-focal answers.

Previous researchers have found that the tendency to give focal answers is associated with lower cognitive ability (Lillard and Willis, 2001; Hurd and McGarry, 1995). In particular, age and education, both strong correlates of mortality risk, are also related to the tendency to give focal answers to probability questions on a wide variety of topics.

2. Probability Beliefs and the Modal Response Hypothesis

In this section, we describe a theoretical model of survey response developed by Lillard and Willis (2001) which attempts to relate answers that an individual gives to a survey question about the subjective probability of a given event and his underlying probability beliefs represented by the density function, $g(p)$. As we discussed in the introduction, Lillard and Willis argue that it is cognitively less burdensome for a respondent to answer a survey probability question quickly by reporting the most likely value of *p*, given by the mode of $g(p)$, than it is to report the expected value given by $\overline{p} = \int pg(p)dp$. Moreover, in this model the mode is often a good approximation of the mean. Although the model is designed to apply to any subjective probability question using the HRS format, in this paper we restrict our discussion to survival probabilities.

Following Lillard and Willis (2001), we assume that person *i*'s subjective beliefs about surviving to a target age can be represented by a heterogeneous probit function

$$p = \Pr(I > 0 \mid \mu_i, \delta) = \Pr(\mu_i + \delta > u) = \Phi(\mu_i + \delta), \tag{1}$$

where *p* is the probability of survival and *I* is an index function,

$$I = \mu_i + \delta - u\ . \tag{2}$$

In this function, μ_i is a parameter that depends on all the information person *i* has about the likelihood of survival.

Uncertainty in this model is represented by the random variable δ in (1) and (2). The special case of risk without uncertainty corresponds to the case in which $\delta = 0$ for sure. In this case, the subjective probability of the event is given by the cumulative standard normal or probit function,

$$p = \Pr(\mu_i > u) = \Phi(\mu_i) \tag{3}$$

The individual is "sure" of his chance of survival and $g(p)$ is degenerate at $\Phi(\mu_i)$. If $\mu_i = 0$, for instance, then $\Phi(0) = 0.5$ and the individual believes he has a fifty-fifty chance of survival for sure. Similarly, if μ_i is one or minus one, he believes his survival risk is 84 percent or 16 percent, respectively.

Increased uncertainty about survival risk, is represented by increased variance in δ. For example, suppose that $\mu_i = 0$ and that the individual believes that δ is equally likely to be plus or minus one so that $g(p \mid \mu_i) = 0.5$ at p=0.84 and p=0.16. The individual's expected probability of survival is still one-half, but unlike the case in which he is sure that $\delta = 0$, the individual's probability beliefs are uncertain. As another example, if $\mu_i = 1$ and δ is equally likely to be plus or minus one, the individual would believe that his survival chance is equally likely to be $\Phi(2) = 0.98$ or $\Phi(0) = 0.50$ and his expected survival probability is $\bar{p} = 0.74$.

Uncertain probability beliefs can be represented quite flexibly by assuming that $\delta \sim N(0, \sigma_{\delta_i}^2)$ where σ_{δ_i} reflects person *i*'s degree of uncertainty, which may vary across people. Since δ is symmetrically distributed about zero under this assumption, it is clear from (1) that the median probability belief—i.e., the median of $G(p \mid \mu_i)$—is equal to $\Phi(\mu_i)$ and that increasing values of σ_δ generate "median-preserving spreads" of *G*. As we discuss later, in many cases a median-preserving spread is also approximately a mean-preserving spread of *G*. If this is the case, an increase in σ_{δ_i}, holding μ_i constant, leads to an increase in uncertainty about p in the same way that a mean-preserving spread of a distribution of payoffs leads to an increase in risk.

Assuming that δ is normally distributed, Lillard and Willis (1978) derive the induced distribution given by

$$G(p) = \Pr(P < p) = \Pr(\Phi(\mu_i + \delta) < p), \tag{4}$$

$$= \Phi\left(\frac{\Phi^{-1}(p) - \mu_i}{\sigma_{\delta_i}}\right). \tag{5}$$

The probability density function is

$$g(p) = \frac{\phi\left(\frac{\Phi^{-1}(p) - \mu_i}{\sigma_{\delta_i}}\right)}{\phi(\Phi^{-1}(p)) \cdot \sigma_{\delta_i}}. \tag{6}$$

Note that *g(p)* only depends on μ_i and σ_{δ_i}.

The mode when *g(p)* is concave and the minimum when *g(p)* is convex are given by

$$p = \Phi\left(\frac{\mu_i}{1 - \sigma_{\delta_i}^2}\right), \tag{7}$$

which is derived by taking the logarithm of *g* and setting its derivative equal to zero. When σ_{δ_i}=0, the distribution is degenerate and the median and the mode are equal. As σ_{δ_i} increases with μ_i constant, the median remains constant, but the mode shifts to the left if μ_i<0 and to the right if μ_i>0. In this way, the difference between the mode and median increases as uncertainty increases.

Figure 2 presents a matrix of 81 probability density functions—$g(p \mid \mu_i, \sigma_{\delta_i})$ in equation (6)—corresponding to nine different values of the median of *G*, given by $\Phi(\mu_i)$ on the horizontal axis, and nine different degrees of uncertainty, measured by σ_{δ_i}, on the vertical axis. In our model, these densities represent different possible beliefs about the probability of survival to a target age. For example, consider the densities for which $\mu_i = 0$ so that the median and mean survival probability is one-half. These are represented by the column of density graphs in the center of the figure. The density in the lowest graph in the column, corresponding to a very low level of uncertainty of $\sigma_{\delta_i} = 0.01$, has almost all its probability mass concentrated at 0.5. As σ_{δ_i} increases, the densities spread. For example, when $\sigma_{\delta_i} = 0.1$, $g(p)$ has a single mode at 0.5 and probability mass near zero for $p < 0.4$ and $p > 0.6$. In this case, an answer of "50" to the survey question means that the respondent believes that the probability is between 40 and 60 percent, so that 50 is a reasonable approximation. When $\sigma_{\delta_i} = 1$, the density is uniform, reflecting a belief that all probabilities between 0 and 1 are equally likely. For values of $\sigma_{\delta_i} > 1$, the density is U-shaped with maxima at 0 and 1. According to the *MRH*, a survey respondent would report "50" in any of these cases. Thus, according to the *MRH*, an answer of "50" by a survey respondent might reflect a very precise view that the chance of an event is one half; a view that the probability is approximately one half; or a view that the probability could be anything so that an answer of "50" indicates epistemic uncertainty.

The effect of increasing uncertainty when the median probability is fairly far from one half can be seen, for instance, by examining the column of densities in Figure 2 for $\mu_i = -1$, corresponding to a median probability of 0.16. The distribution begins to spread out as the level of uncertainty increases, but has a single mode near 15 percent. According to the *MRH*, a survey respondent would provide an answer of "15" in this case. When σ_δ reaches a level of about 0.75, the density function becomes J-shaped with a mode at zero and the *MRH* implies that a survey respondent would report "0". As σ_δ continues to increase and reaches a value of about 5 or more, $g(p)$ becomes U-shaped with maxima near 0 and 1 and, in this situation of epistemic uncertainty, the *MRH* implies an answer of "50".

Similarly, according to the *MRH*, a person with $\mu_i = 1$ and a median probability of 0.84 would give a response of 84 if he had a low degree of uncertainty; a value of "100" when his uncertainty is large enough to give $g(p)$ a J-shape with a mode at 1; and a value of "50" for very high degrees of uncertainty which generate a bimodal $g(p)$ with modes a 0 and 1.

The matrix of densities in Figure 2 is meant to illustrate the possible beliefs that a given person might have concerning his subjective probability of survival to a target date. The shaded areas in the figure divide this space into four regions corresponding to the four types of answers implied by the *MRH*. Exact answers corresponding to unimodal distributions with modes strictly greater than zero and less than one occur in the inverted U-shaped area for persons with low levels of uncertainty. Note that it requires a fairly low degree of uncertainty—or, in other words, fairly precise probability beliefs—in order for a respondent to give an exact answer to very low or high probability events. With more uncertainty—say, $\sigma_{\delta_i} = 0.75$— persons with the same value of μ_i would give answers of "0" or "100". On the other hand, a survey respondent with that level of uncertainty would distinguish between chances of 30 and 40 percent corresponding, respectively, to median probabilities $\Phi(-0.5)$ and $\Phi(-0.25)$.

The other three shaded regions of the diagram correspond to the different types of focal answers. The U-shaped border of the region containing answers of "50" implies that uncertainty can dominate the response at lower levels of uncertainty when the median probability is closer to "50". The pie-shaped regions associated with focal answers of "0" and "100" indicate, not surprisingly, that a lower level of uncertainty is compatible with these focal answers for low and high probability events, respectively, than for situations in which the median probability is closer to one half.

Although we treat the *MRH* as a maintained hypothesis in this paper, it is worth noting that Lillard and Willis (2001) use data on responses to the HRS question about the probability that tomorrow will be sunny to test whether focal and non-focal answers to this question follow the patterns shown in Figure 2. To do so, they exploit the fact that the HRS survey is conducted using households in a stratified national sample of primary sampling units (PSUs) with a field period covering all seasons of the year. For purposes of the test, in each of three survey waves of they divide the sample into cells containing households who lived in the same PSU and were surveyed in the same month. Within each PSU-month cell they calculate (a) the mean probability report of persons in each cell and (b) the fraction of responses that were of each type: "exact" (i.e., non-focal), "0", "50" and "100." They assume that variation in actual weather across cells is reasonably estimated by the mean probability of a sunny day and, further, that it is plausible to assume that weather is exogenous so there is no correlation between the mean probability and the (unobservable) average level of uncertainty across cells.

Under these assumptions, (smoothed) plots of a given type of response, measured on the vertical axis, versus average probability of a sunny day measured on the horizontal axis should follow the patterns depicted in Figure 2. Using a non-parametric smoother, Lillard and Willis (2001) find these plots to be in almost perfect accord with the predictions of the *MRH* in all three survey waves. Specifically, the fraction of non-focal answers has an inverted U-shape, rising to a peak at $p = 50$ and then falling; the fraction of answers at "0" and "100" are, respectively, monotonically decreasing and increasing functions of *p*; finally, the fraction of "50" answers also follows an inverted U-shaped pattern with a maximum at $p = 50$.

While these results provide empirical support for the theory of survey response underlying the *MRH*, it is not plausible to assume that the level of survival risk and the degree of uncertainty about that risk are uncorrelated. For example, other things equal, better educated people have higher life expectancy and may also have better knowledge of actuarial risks for their demographic group and a better understanding of the mortality risks associated with particular diseases or life style behaviors. Our econometric model will allow for such correlations.

3. Likelihood Function for Modal Response Hypothesis

The modal response hypothesis forms the basis of the econometric model that we use in this paper to estimate the determinants of probability beliefs from survey answers to subjective probability questions that use the HRS format. Suppressing the *i* subscript for simplicity, we assume that the parameter μ in (1), which determines the subjective median survival probability, may be written as

$$\mu = Z_p^* \beta_p^* , \tag{8}$$

where Z_p^* is a vector of variables that includes all the information the respondent has about the likelihood of the relevant outcome and β_p^*, represent the weights placed by the individual on each of these pieces of information in forming his expectations. Similarly, we assume that an individual's subjective uncertainty given by the parameter, σ_δ, in (1) may be written as

$$\ln \sigma_\delta = Z_a^* \beta_a^* \quad (9)$$

where Z_a^* is a vector of all variables that influence the individuals degree of uncertainty or ambiguity about the survival risk.

In practice, the econometrician observes only a subset of the factors that determine an individual's beliefs. In addition, we do not have sufficient data to identify the differences across persons in the weight given to each observed factor in forming their beliefs. Hence, in our econometric model, we assume that we observe a subset of determinants of beliefs given by the vector, (Z_p, Z_a), and also assume that all individuals weight these determinants in the same way according to the vector (β_p, β_a) which does not vary across persons. Given these assumptions, we may write the empirical counterparts to (8) and (9) as

$$\mu = Z_p \beta_p + u_p \quad (10)$$

and

$$\ln \sigma_\delta = Z_a \beta_a + u_a \quad (11)$$

where the random error terms capture the effects of unobserved components of (Z_p^*, Z_a^*). They are assumed to follow a correlated bivariate standard normal distribution, $(u_p, u_a) \sim N(0,0,1,1,\rho)$, where the parameter, ρ, measures correlation between the unobserved determinants of risk and uncertainty. The variances of u_p and u_a are not identified and they are therefore normalized to unity. We also assume that $E(u_p Z_p) = E(u_a Z_a) = 0$. Finally, conventional exclusion restrictions are needed for identification. Thus, we assume that some elements of Z_p do not appear in Z_a and conversely. We discuss the substantive aspects of this model and our identifying assumptions as applied to subjective survival probabilities below in Section 4.

The goal of our econometric model is to estimate the subjective survival probability beliefs of individuals, $G(p \mid Z_p^*, Z_a^*)$, from survey responses to subjective probability questions under the modal response hypothesis. This function is obtained by substituting (9) and (10) into (4) and setting $u_p = u_a = 0$. We will use a diagrammatic exposition to develop the intuition underlying the likelihood function for this model; the formal mathematical derivation can be found in Appendix I.

In Section 2, we explained the *MRH* with the aid of a matrix of 81 graphs of density functions, $g(p \mid \mu, \sigma_\delta)$, portrayed in Figure 2. Recall that the figure is partitioned into four regions corresponding to four types of survey responses: "exact", "0', "50" and "100". According to the *MRH*, the type of response given by a particular individual is determined by the region into which the point (μ, σ_δ) falls.

Consider, for example, a group of observationally identical individuals with $(\mu, \sigma_\delta) = (0,1)$ corresponding to the uniform distribution in middle graph in Figure 2. The actual beliefs of any individual *i* in this group are determined by the parameters $(\mu_i, \sigma_{\delta_i}) = (u_{p_i}, e^{u_{a_i}})$ where u_{p_i} and u_{a_i} represent components of survival risk and subjective uncertainty that are known to the individual but are not observed by the econometrician. Among a group of such individuals, different combinations of (u_{p_i}, u_{a_i}) could generate all four types of response with "exact" and "50" responses more likely, respectively, for persons with negative or positive values of u_{a_i} for any given value of u_{p_i}. Similarly, responses of "0" and "100" are more likely, respectively, for persons with negative or positive values of u_{p_i} for any given value of u_{a_i}.

Given the assumption that u_{p_i} and u_{a_i} are jointly normally distributed and further assuming $\rho = 0$, for groups of individuals with $(\mu, \sigma_\delta) = (0,1)$ we would expect to see equal fractions of "exact" and "50" responses and equal fractions of "0" and "100" responses. In addition, we would expect see the values of the "exact" responses distributed symmetrically about one half. Conditional on the parameters of the model (i.e., β_p, β_a and ρ), the likelihood of each type of answer and the distribution of "exact" answers will vary in predictable ways for persons with different observed characteristics given by Z_p and Z_a.

This is the intuition underlying the likelihood function for the *MRH* that we estimate in Section 5. A more detailed derivation of the likelihood function is presented in Appendix 1 and a diagrammatic representation is given in Figure 3. This figure transforms the axes of Figure 2 so that $(\mu_i, \ln \sigma_{\delta_i}) = (Z_{p_i}\beta_p + u_{p_i}, Z_{a_i}\beta_a + u_{ai})$ can be plotted on an arithmetic scale and the bivariate normal distribution of (u_{p_i}, u_{a_i}) can be represented by elliptic contours centered on $(Z_{p_i}\beta_p, Z_{a_i}\beta_a)$. Contours in Figure 3 are drawn for the case in which the mode of the subjective distribution is 0.65. This space is portioned into four regions by two functions, labeled *h* and *k*, which trace out the boundaries of the four-regions (this is covered more fully in Appendix 1). Points in the figure that lie below both curves correspond to "exact" answers; points above both curves correspond to answers at "50". Points lying between the two curves correspond to answers of "0" or "100". The likelihood of each of these types of answers depends on the fraction of the bivariate normal that falls into each region. Conditional on being in the exact region, the likelihood of a given exact answer such as "65" is given by the area of the bivariate normal that falls within the region marked A in Figure 3.

4. Empirical Analysis

Beliefs about subjective survival probabilities presumably depend, on an individual's knowledge of his situation, his ability to translate this information into a probability and on his level of optimism or pessimism. In the empirical model estimated in this section we use the information available in the HRS to try to capture several of the major determinants of beliefs in a parsimonious fashion.

4.1 Sample and Measures Employed

The sample used in the empirical analysis consists of 13,505 respondents to the 2002 Health and Retirement Study, over age 50 in 2002, who provided responses to the subjective probability of survival questions. Excluded from the sample are proxy respondents and nonrespondents in the 2000 wave of data collection. Also excluded are persons over 90 who were not asked the survival probability questions. Table 1 presents the statistics for the variables used in our analyses. The average age of our sample members is just over 68 years (ranging from 51 to 90 years) and on average the target age was 14.1 years from their current age. The modal sample member is a white female with a high school education although there is substantial variance in each of these dimensions.

[Table 1 about here]

Self-rated health in the HRS is measured on a five-point scale--1) Excellent, 2) Very Good, 3) Good, 4) Fair and 5) Poor. We translated these into three categories: 1) excellent/very good; 2) good; 3) fair/poor. We then constructed dummy variables representing the combination of self-rated health in 2000 and 2002 for each respondent with excellent/very good in both years as the baseline case. Table 1 shows that mean self-rated health in both years is very similar, with a small worsening of health in 2002.

Two cognitive measures are used in the analyses—total recall and "serial 7s". Total recall is a measure of the individual's short-term memory. Respondents are read a list of ten words and then asked to recall and repeat them. This is done both immediately after reading the list and then again a few minutes later. The best possible score, therefore, is twenty and the average respondent actually scores slightly more than 10. "Serial 7s" is a measure of cognitive processing capacity and attention. Respondents are asked to recursively subtract 7 beginning at 100. The interviewer stops the respondent after five such subtractions and records the number of correct reports. Thus, the maximum score is five and the average is three and two-thirds.

Since mood can affect the level of optimism we also include in our model a count of the number of depressive symptoms the respondent has exhibited over the recent past. These range from disturbed sleep patterns, through feelings of hopelessness all the way up to thoughts of suicide. In all there are eight such symptoms measured and they are used to construct the CESD depression scale (Ofstedal, et. al., 2002). On average respondents have less than one and one-half depressive symptoms.

The final two measures included in our analyses are derived from responses to subjective probability questions other than survival—"optimism" and "focal propensity". Following Kezdi and Willis (2003) the optimism index is constructed from factor analysis of normalized scores on all subjective probabilities other than survival with positive loadings for the probabilities of favorable events (e.g. income keeping up with inflation) and negative loadings for unfavorable events (e.g. major economic depression in the next ten years). Since it is based on standardized scores, the mean has little substantive meaning. Focal propensity, used an a measure of uncertainty by Lillard and Willis (2001) and Kezdi and Willis (2003), is simply the percent of all subjective probabilities asked (other than the survival probabilities) which are answered either 0, 50, or 100.

4.2 Maximum Likelihood Model Estimates

Table 2 presents the maximum likelihood parameter estimates of the modal response model for three specifications. The model was estimated via a modified version of the Davidson-Fletcher-Powell algorithm with function evaluations using numerical integration of the bivariate normal pdf and Gaussian quadratures. Numerical integration is required because the likelihood function contains CDFs for the various non-rectangular regions described in Figure 3 above. In each of these regions integration involves 100 evaluations per observation of the PDF and thus, estimation is numerically intensive—Model 3 in the table, for instance, required more than 1.5 billion calls to the bivariate normal PDF and individual function evaluations to achieve convergence. In all three models convergence of the likelihood function and the parameter vector was achieved and the gradient vanished. Given the numerical intensity of the model, we must use parsimonious specifications.

Age enters the survival-index portion of the model in a complex and non-linear fashion—specifically we use a variant of the Gompertz survival function:

$$S(Age,T) = \exp[\frac{-\exp(\gamma Age)}{\gamma}[\exp(\gamma(T - Age) - 1]]$$

where T is the "target age" used in the subjective survival question and γ is the Gompertz parameter to be estimated. To express this in index units we take its inverse normal. Thus, the index becomes:

$$Z_p\beta_p + \Phi^{-1}(S(Age,T))$$

and the β_p's become interpretable as deviations from the expected survival associated with the Z_p's (see Appendix 2 for details).

For separate identification of the parameters associated with the level of risk with the degree of uncertainty, as noted earlier, we need to specify some variables in Z_p that affect the median probability in (9) but do not affect uncertainty and, conversely, some variables in Z_a that affect uncertainty in (10) but are assumed not to affect survival risk. In Model 1, the first specification presented in Table 2, identification is achieved by having age and target date enter the survival-index equation through the Gompertz functional form while it enters linearly in the focal equation which also includes the cognitive measures. Age, in the Gompertz form, is highly significant as are gender, race and education. Males have lower survival expectations than females and the more highly educated have higher expectations. These effects are consistent with the results of most other studies of actual survival. The race effect, while consistent with earlier more straight-forward analyses using HRS data (Hurd and McGarry, 1995), is not consistent with life-table estimates. This may reflect race-mortality cross-over found in actuarial tables, it may mean that non-whites are more optimistic than whites or it may reflect response error associated with respondent interpretation of the questions.

To gain an intuitive feel for the meaning of the estimates of γ in Table 2, we simulated the model with the $Z_p\beta_p$ and the respondent's actual age and then again with his age augmented by 1 year in the index portion of the model. Figure 4 compares the actual and predicted probabilities for the base case and although the predicted values are a little low for the 0 and 100 predictions and a bit high for the 50, overall the distributions are quite similar. For Model 1, augmenting age by one year in the survival index portion of the model while holding the horizon (T-Age) constant, results in a 2.13 percentage point decline in average subjective survival from 59.97 percent to 57.83. Thus a positive γ implies a negative association of age and subjective survival probabilities. This relationship is consistent with that found by Hurd and McGarry (1995). Similarly, a one-year increase in the target age (holding actual age constant) results in a 4.92 percentage point decline. This is a larger decline than that associated with age because of compounding.

As noted above, the other coefficients in the Survival Index portion of Models 1 and 2 are also consistent with the Hurd-McGarry estimates. One thing not included in the Hurd-McGarry model is the number of depressive symptoms as measured by the CESD index which is negatively associated with subjective survival chances.

In the focal, or uncertainty, portion of the model, age has a highly significant, positive, but weak, effect on the propensity to provide focal answers and education has a significant (though small) negative effect. This suggests that uncertainty increases with age but is reduced by education. When we simulate the effect of a one-decade increase in age (in the focal portion of the model only) we find that the overall effect is a modest 2.09 percentage point increase in focal responses. Most (1.58 percentage points) of this increase is in epistemic uncertainty and most (0.50 percentage points) of the rest appears in the form of increased focal responses at 0.

These findings are also consistent with older and less educated respondents having more difficulty providing carefully considered responses and relying instead on quick and easy heuristics resulting in focal answers. Each of the models, however, controls for cognitive ability via the Total Recall and Serial 7s measures. Because these enter both additively and interactively, it is somewhat difficult to interpret their overall effect but we can simulate the relationships as we did for age and horizon above. When we do so we find that, despite their statistical significance, the effect of cognition on the percent of focal answers is quite small. A one unit (10%) increase in short-term memory reduces the estimated proportion of focal answers by only 0.13%. Similarly, a one-unit (nearly 30%) increase in the Serial 7s score reduces the estimated proportion of focal answers by only 0.29%.

[table 2 about here]

In Model 2 the optimism index is added to the survival index and the focal propensity measure to the focal index. The coefficients on both variables are large, positive and highly significant. The major impact of this on the other parameter estimates in the survival function is to increase the estimated effect of race and gender while reducing (to insignificance) the estimated effect of education. In the focal portion of the model, inclusion of the focal propensity measure has the effect of totally wiping out the strong positive effect of age. Evidently older respondents tend to provide focal answers not only for survival probabilities but for other probability questions as well.

Finally, in Model 3 dummy variables indicating self-rated health combinations in 2000 and 2002 are added to both portions of the model. Subjective Survival probabilities increase with self-rated health in both 2000 and 2002. Additionally, higher self-rated health in either year leads to reduced uncertainty, but this effect is somewhat inconsistent across all the dummy variables. Interestingly, adding health to the survival index substantially reduces the impact of depression as measured by the CESD. It is hard to say, however, if poor health is a contributor to depression or depression causes respondents to provide more pessimistic responses

Overall the modal response model provides estimates of associations of subjective risk and various predictors that are consistent with those found elsewhere in the literature. Age, gender, race, education, optimism and health are all highly significant predictors of subjective risk. Additionally, the modal response model provides estimates of the effects of various factors on uncertainty that, while reasonable in sign, are not as powerful as they are in the survival risk portion of the model.

4.3 How Uncertain are HRS Respondents?

The model implies that there is a great deal of subjective ambiguity or uncertainty about mortality risks. Figure 5 presents the estimated subjective probability density functions implied by the coefficients in Table 2 for three levels of the focal propensity. The left-most distribution in the figure corresponds to a focal propensity at one standard deviation above the mean, the middle distribution has focal propensity at the mean and the right most distribution is one standard deviation below the mean. These correspond, respectively, to imprecise, average and precise beliefs. In all three cases there is substantial spread suggesting great uncertainty—even for “precise” beliefs. The variance of δ declines from 0.76 to 0.59 as the precision of beliefs increases from low to high, but even at 0.59 the cumulative subjective probability that actual survival chances are less than one half the mode is 7.3%. Thus, the odds that the true survival probability is less than one-half the mode are better than one in thirteen, even for respondents with very precise beliefs.

Another important thing to note about Figure 5 is that, while constructed to exhibit a median-preserving spread, it is also approximately mean-preserving—the means for the three levels of precision are 0.57, 0.58, and 0.57, respectively. This being the case, the arguments presented in Section 2, above, imply that an increase in σ_δ^2 leads to increased uncertainty.

Figure 6 shows the implied distribution for optimistic (those with an optimism index one standard deviation above the mean), average and pessimistic respondents. Optimism enters the model only through the survival index. The mode declines monotonically from 0.80 for optimists to 0.46 for pessimists. Both the mean and median also decline but not as dramatically. Since the optimism index does not enter the focal portion of the model, $\sigma_\delta^2 = 0.62$ for all three plots. Nevertheless, the spread of the distribution is affected by changes in the optimism index. The standard deviation of the belief distribution for the optimist (0.19) is slightly less than that of the average individual (0.20) and less than that of the pessimist (0.21).

Finally, when variables enter both the survival and focal indices, as does self-rated health, all the moments of the distribution are affected. Figure 7 presents the belief distributions for nine combinations of Self-rated Health in 2000 and 2002. The upper-middle and upper-right plots show the effects of health declines from "excellent/very good" in 2000 to "good" and "fair/poor" in 2002. As one moves from left to right in the top row the mode drops from 0.64 to 0.33 and the σ_{δ}^{2} (not shown) increases from 0.62 to 0.68.

The graphs in Figure 7 can be used to illustrate some potentially important points about the difference between modal responses and expected survival risk. Consider the graphs on the diagonal which present estimated density functions for a typical HRS respondent who reported the same subjective health state—excellent/very good, good, or fair/poor—in both 2000 and 2002. Note that all the graphs are unimodal so that, according the *MRH*, the person would report an exact answer to a survey question about their chance of living to the target age. If the person is in excellent/very good health, the upper left graph implies, according to the *MRH*, that he would respond with "64" when asked the HRS survival probability question because the mode of the estimated distribution is 64. Similarly, a person whose health is good in both years would answer with a 37 percent chance of survival to the target age and someone who was in fair/poor in the right-most graph in the third row would report only a 10 percent survival chance.

Suppose that respondents have rational expectations in the sense of an affirmative answer to the question posed by Smith, et. al. (2001): "Can People Predict Their Own Demise?" Also assume that the estimated density functions in the diagonal of Figure 7 accurately capture their expectations. If we follow these people until the target date is reached, the fraction we expect to remain alive is given by the mean of the relevant density. Thus, 58 percent of those who were in excellent/very good health in both 2000 and 2002 would be expected to survive as compared to the 64 percent probability of survival they reported on the survey in 2002. At the other extreme, those in fair/poor health told us they had only a 10 percent chance of surviving to the target date, but the estimated density implies that we should expect 28 percent to survive. Some researchers might accuse those in excellent/very good health of over-optimism and those in fair/poor health of over-pessimism; other researchers might suggest that their probability reports contain measurement error. While there is some merit in both of these views, even for our example with assumed rational expectations, our own interpretation is that the single number elicited in the HRS probability question is not sufficient by itself to be used to produce good predictions of outcomes, especially when risks are very high or very low. Either we need to find ways to improve the elicitation of subjective probabilities on surveys or we need to recalibrate the answers to existing questions using econometric models such as the one presented in this paper.

5. Conclusions

The modal response hypothesis is used in this paper as the foundation for an econometric model that is intended to provide a mapping between survey responses to probability questions and the underlying subjective probability beliefs of individuals about their chances of surviving to a target age. In this paper, we have presented the *MRH* as a hypothesis designed to capture the kinds of "gut response" to such questions that would be made after about 15 seconds of consideration by persons who vary the amount of information they have about actuarial risks to health, about their own health-related circumstances and in their capacity process such information into subjective beliefs. We argued in Section 2 that reporting the mode is relatively easier from a cognitive point of view than the mean or the median; that the mode often provides a good approximation to the expected probability that is called for in *SEU* theory; and that the hypothesis is consistent with the observed tendency of many survey respondents to give focal answers. Under the *MRH*, a high level of subjective uncertainty leads to a higher propensity to give focal answers and, under certain conditions, increased uncertainty leads to more risk-averse behavior. This latter hypothesis has received support in empirical research on savings, portfolio choice and stock ownership decisions by Lillard and Willis (2001) and Kezdi and Willis (2003) who find significant effects of the propensity to give focal answers on such decisions.

A reader may justly wonder how we can advance the *MRH* as a "fast and frugal" algorithm for boundedly rational, unsophisticated respondents in view of the fairly complex mathematical manipulations needed to derive the distribution of subjective beliefs, $g(p)$ in (6). In part, our own methodological position goes back to the classic "as if" defense of "unrealistic" assumptions in economic theory provided by Milton Friedman's example of an expert billiard player. A great billiard player does not line up his shots by doing calculations based on the laws of physics. However, to a scientist observing his behavior, it will appear that he is acting as if he were making such calculations. To economists, individuals are like expert billiard players, successfully solving complex problems using experience and instinct.

In our case, we assume that respondents have some kind of mental images of the central tendencies of the risks they face and the possible range of these risks. Our representation of these images is given by the patterns of subjective densities portrayed in Figure 2. Since people deal with uncertainty of all sorts in their everyday lives, we assume that the brain has some way of recognizing these patterns but we are ignorant about how it accomplishes this. Finally, we recognize that the "as if" methodology is a way station, hopefully on the path to a fuller understanding to an understanding of how billiard players play pool without knowledge of physics and how consumers assess risk and uncertainty without knowledge of probability theory. Recent research using psychological experiments (Schwarz, 2002) or economic experiments in combination with neuro-imaging (Glimcher and Rustichini, 2004) to study the roles of mood and cognition in judgment show promise in this respect.

Our empirical findings suggest that there is considerable heterogeneity in subjective survival risks, some of it associated with standard demographic variables such as age, sex, race and education. But much of this heterogeneity is related to a general optimism/pessimism factor that is exhibited by respondents on other HRS probability questions, to depressed mood, and to personal circumstances indicated by personal health status and heath shocks. There is also substantial uncertainty about these risks which is manifested by considerable spread in the estimated distribution of subjective survival probabilities for a typical respondent. In addition, we find significant variation in uncertainty, holding expected survival risk constant. Those with better education and higher cognitive scores have less uncertainty; those in better health have more precise beliefs. Beyond these factors, we find that the propensity to give focal answers to other HRS questions has a substantial effect on uncertainty about survival, suggesting that unmeasured factors that increase or reduce a given person's uncertainty have effects that cut across many different domains. It remains for future work to explore the explanation of these findings more deeply and to see whether survival risk and uncertainty about this risk play a role in decisions made by HRS respondents.

References

Bruine de Bruin, W., B. Fischhoff, B. Halpern-Felsher, and S. Millstein (2000), "Expressing Epistemic Uncertainty: It's a Fifty-Fifty Chance," *Organizational Behavior and Human Decision Processes*, 81, 115-131.

Christakis, Nicolas A. (1999) Death Fortold: Prophecy and Prognosis in Medicare Care. University of Chicago Press.

Fischhoff, B . and W. Bruine de Bruin (1999), "Fifty-Fifty = 50%?" *Journal of Behavioral Decision Making*, 12, 149-163.

Gan, L., Hurd, M., and McFadden, D., "Individual Subjective Survival Curves 2003." National Bureau of Economic Research. NBER Working Paper 6478. Forthcoming in

Horn, J.L. (1991). Measurement of intellectual capabilities: A review of theory. Chapter 7 in McGrew, K. S., Werder, J. K. & Woodcock, R. W. *Woodcock-Johnson Technical Manual* (pp. 197-246). Allen, TX: DLM Teaching Resources.

Hurd, M. D. and McGarry, K., The Predictive Validity of Subjective Probabilities of Survival The Economic Journal, vol. 112, 2002. [2002] (Journal Article)

Hurd, M.D. and Smith, J.P., Expected Bequests and their Distribution 2002. RAND. [2002] (Monograph)

Gan, Li, Michael Hurd and Daniel McFadden (2003), "Individual Subjective Survival Curves," National Bureau of Economic Research Working Papger 9480. Forthcoming in David Wise, ed. *Analyses in the Economics of Aging* . University of Chicago Press.

Glimcher, Paul W. and Rustichini, Aldo. (2004) "Neuroeconomics: The Consilience of Brain and Decision." *Science* vol. 306, no. 15, pp. 447-52.

Hurd, M. and K. McGarry (1995), "Evaluation of the Subjective Probabilities of Survival in the Health and Retirement Study," *Journal of Human Resources*, 30, S268-S292.

Hurd, M. and K. McGarry (2002), "The Predictive Validity of Subjective Probabilities of Survival," *The Economic Journal*, 112, 966-985.

Hurd, Michael D. and James P. Smith (2001), "Anticipated and Actual Bequests," in *Themes in the Economics of Aging*, David Wise, ed. Chicago: University of Chicago Press, pp. 357-389.

Hurd, M. and J. Smith (2002), Expected Bequests and Their Distribution," National Bureau of Economic Research Working Paper 9142.

Kezdi, Gabor and Robert J. Willis (2003), "Who Becomes A Stockholder? Expectations, Subjective Uncertainty, and Asset Allocation," presented at Annual Conference of the Retirement Research Consortium, Washington, D.C., May 15-16, 2003.

Knight, Frank H. (1921) *Risk, Uncertainty, and Profit*. Boston, MA: Hart, Schaffner & Marx; Houghton Mifflin Company.

Lillard, Lee A. and R. J. Willis (1978), "Dynamic Aspects of Earnings Mobility," *Econometrica*, 46, 985-1012.

Lillard, Lee A. and R. J. Willis (2001), "Cognition and Wealth: The Importance of Probabilistic Thinking," Michigan Retirement Research Center Working Paper UM00-04.

Manski, C., (2004) "Measuring Expectations," *Econometrica*.

McArdle, J.J., Ferrer-Caja, E., Hamagami, F. & Woodcock, R.W. (2002). Comparative longitudinal multilevel structural analyses of the growth and decline of multiple intellectual abilities over the life-span. *Developmental Psychology, 38* (1) 115-142.

Ofstedal, Mary Beth; Fisher McAuley, Gwenith Fisher McAuley and Herzog, A. Regula. 2002, "Documentation of Cognitive Functioning measures in the Health and Retirement Study. HRS/AHEAD Documentation Report DR-006 (2001-02)". Report Last Updated: 08/26/02.

Schwarz, Norbert (2002) "Situated Cognition and the Wisdom of Feelings: Cognitive Tuning." In L. Feldman Barrett and P. Salovey, (eds.), *The Wisdom in Feeling*, New York: Guildford Press, pp. 144-166.

Sloan, F. A., Smith, V. K., and Taylor, D. H.(2003) *Parsing the Smoking Puzzle: Information, Risk Perception, and Choice*, Cambridge: Harvard University Press, 2003.

Smith, V. K., D. H. Taylor Jr., F. A. Sloan, F. Reed Johnson, and W. H. Desvousges (2001), "Do Smokers Respond to Health Shocks?," *The Review of Economics and Statistics*, 83, 675-687.

Smith, V. K., Taylor, D. H., and Sloan, F. A. (2001) Longevity Expectations and Death: Can People Predict Their Own Demise? *The American Economic Review*, vol. 91, pp. 1126-1134.

Starmer, Chris. (2000). "Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk" Journal *of Economic Literature*, Vol. 38, No. 2, pp. 332-382.

Willis, Robert J. (forthcoming) "Discussion of Li Gan, Michael Hurd and Daniel McFadden, 'Individual Subjective Survival Curves.'" In David Wise, ed. *Analyses in the Economics of Aging* . University of Chicago Press.

Table 1: Descriptive Statistics

	Mean	Standard Deviation
Subjective Survival Probability	55.23	32.32
Age	68.20	8.94
Horizon	14.10	2.67
Male	0.40	0.49
White	0.85	0.36
Education	12.55	3.00
Optimism	0.03	0.86
Self Rated Health 2000	1.75	0.79
Self Rated Health 2002	1.82	0.80
Total Recall	10.11	3.58
Serial 7s	3.66	1.61
Recall x Serial 7s	38.97	22.62
%Heaped Qs	0.66	0.24
CESD	1.44	1.93

Table 2: Estimates of Modal Response Model

Subjective Probability of Survival to Age X
Modal Response Model (Standard Errors in Parentheses)
2002 HRS--13,505 Individuals over 50 Years of Age
**=significant at 99% *=significant at 95% +=significant at 90%

	Model 1		Model 2		Model 3	
F(ρ)	0.343**	(0.042)	0.325**	(0.048)	0.400**	(0.052)
ρ	0.145**		0.136**		0.173**	
Survival Index:						
Survival Constant	2.145**	(0.016)	2.339**	(0.073)	2.813**	(0.061)
Gompertz γ	0.042**	(0.0001)	0.040**	(0.0004)	0.040**	(0.006)
Male	-0.184**	(0.020)	-0.246**	(0.020)	-0.203**	(0.020)
White	-0.266**	(0.028)	-0.371**	(0.030)	-0.437**	(0.029)
Education	0.035**	(0.003)	0.006	(0.004)	-0.101**	(0.039)
CESD	-0.124**	(0.003)	-0.106**	(0.005)	-0.050**	(0.006)
Optimism	NA		0.322**	(0.013)	0.244**	(0.013)
2000 Health; 2002 Health:						
Exc./v. good; good	NA		NA		-0.095*	(0.044)
Exc./v. good; fair/poor	NA		NA		-0.465**	(0.097)
Good; exc./v. good	NA		NA		-0.237**	(0.039)
Good; good	NA		NA		-0.373**	(0.033)
Good; fair/poor	NA		NA		-0.507**	(0.055)
Fair/poor; exc./v. good	NA		NA		-0.459**	(0.069)
Fair/poor; good	NA		NA		-0.634**	(0.050)
Fair/poor; fair/poor	NA		NA		-0.927**	(0.037)
Focal Index:						
Focal Constant	-0.441**	(0.063)	-0.850**	(0.110)	-0.913**	(0.115)
Age (decades)	0.047**	(0.006)	-0.005	(0.011)	-0.001	(0.001)
Male	-0.025	(0.020)	-0.016	(0.021)	-0.021	(0.021)
Education	-0.035**	(0.004)	-0.025**	(0.004)	-0.226**	(0.039)
Total Recall	0.050**	(0.017)	0.056**	(0.018)	0.050**	(0.019)
Serial 7s	0.013+	(0.007)	0.015*	(0.007)	0.014+	(0.008)
Recall x Serial 7s	-0.005**	(0.002)	-0.005**	(0.002)	-0.004*	(0.002)
CESD	0.002	(0.006)	-0.004	(0.005)	-0.009	(0.006)
Focal Propensity	NA		0.880**	(0.049)	0.850**	(0.056)
2000 Health; 2002 Health:						
Exc./v. good; good	NA		NA		0.103*	(0.043)
Exc./v. good; fair/poor	NA		NA		0.103	(0.090)
Good; exc./v. good	NA		NA		0.144**	(0.037)
Good; good	NA		NA		0.211**	(0.032)
Good; fair/poor	NA		NA		0.214**	(0.050)
Fair/poor; exc./v. good	NA		NA		0.073	(0.070)
Fair/poor; good	NA		NA		0.175**	(0.048)
Fair/poor; fair/poor	NA		NA		0.092**	(0.035)
Log-Likelihood	-47,050.610		-46,635.273		-46,272.731	

Figure 1. Distribution of Survival Probabilities to Target Age by Age of Respondent

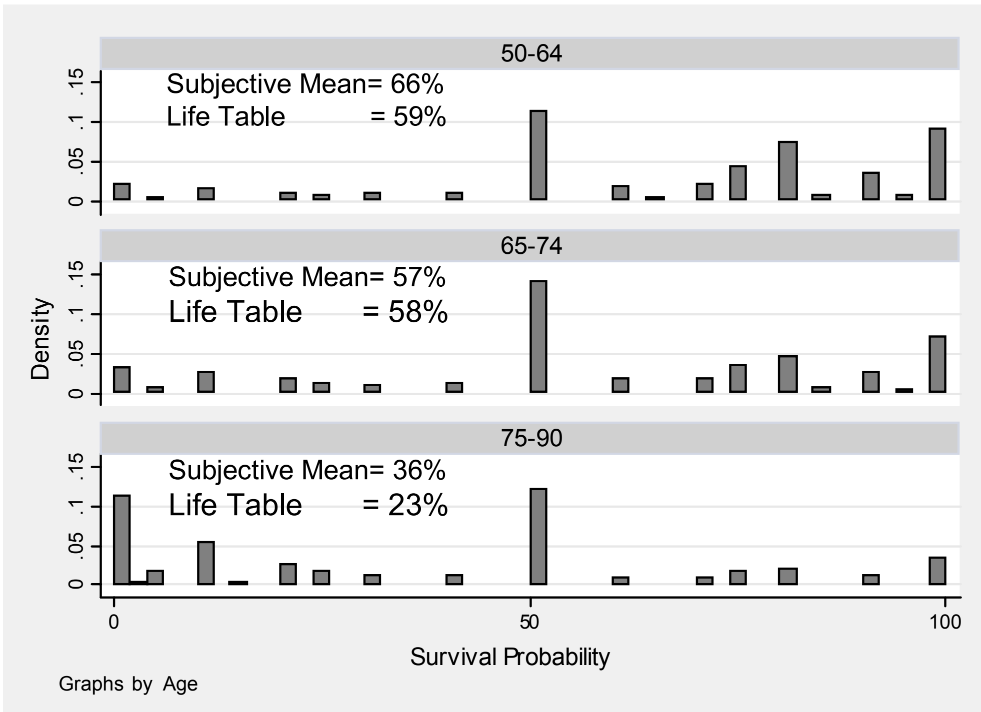


Figure 2: Modal Response Hypothesis: Prior Beliefs and Survey Response

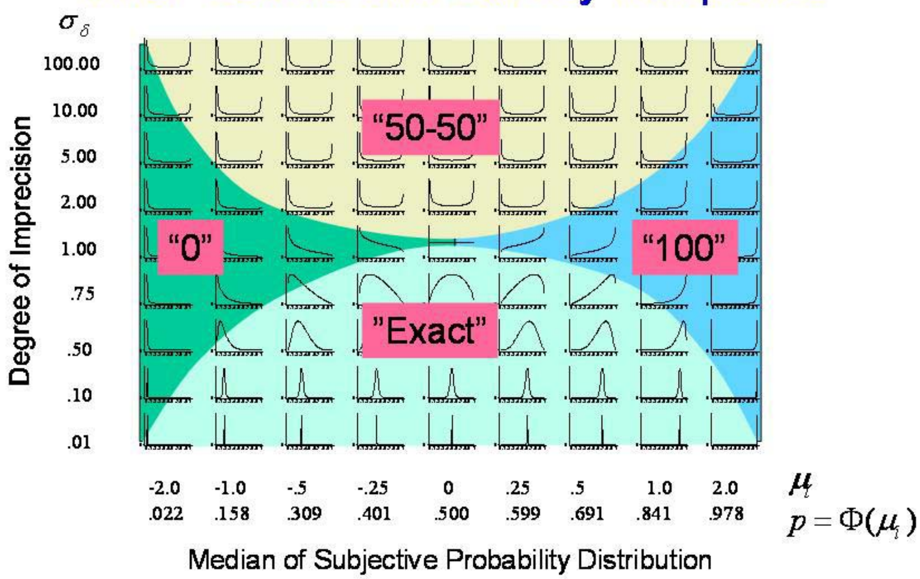


Figure 3. Diagrammatic Representation of the Likelihood Function for Modal Response Hypothesis

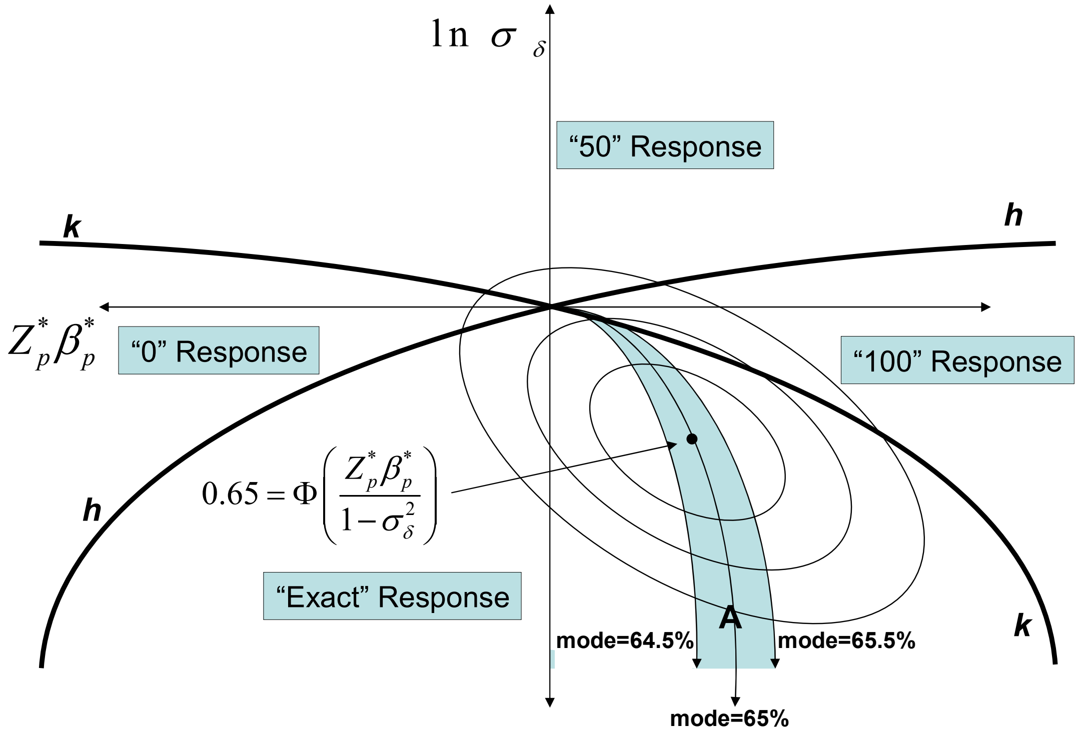


Figure 4. Distribution of Subjective Probabilities of Survival to Age X

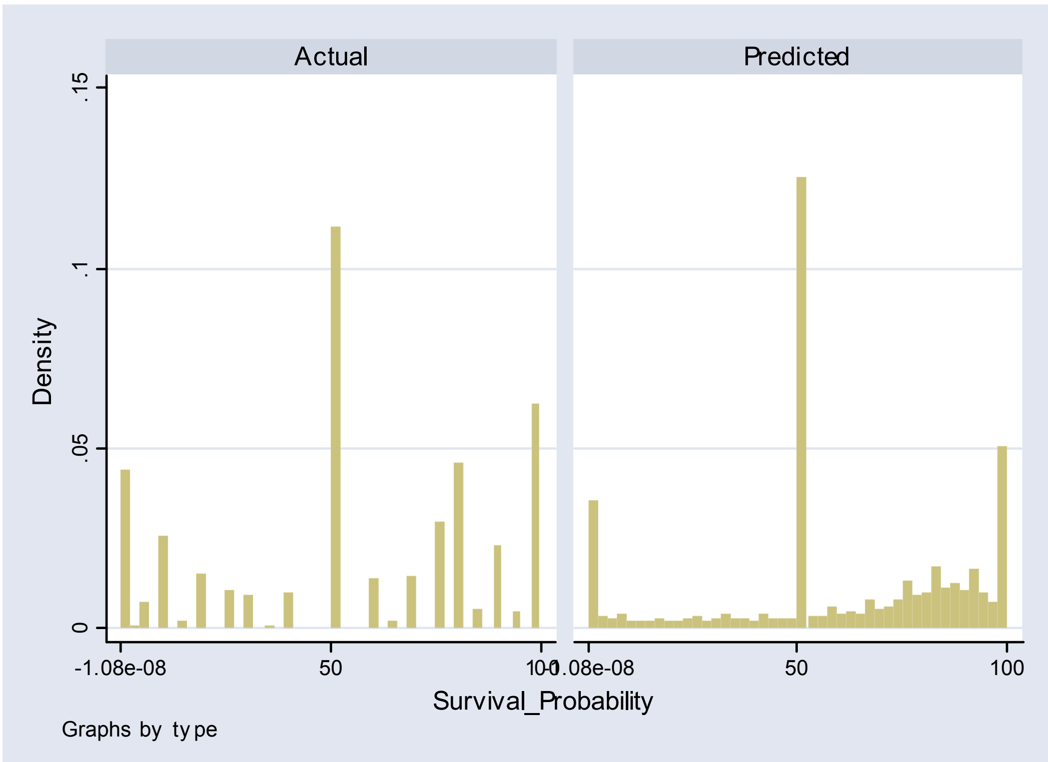
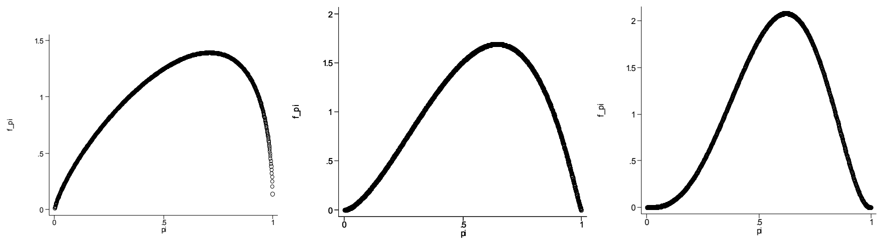


FIGURE 5: THE EFFECT OF A CHANGE IN FOCAL PROPENSITY

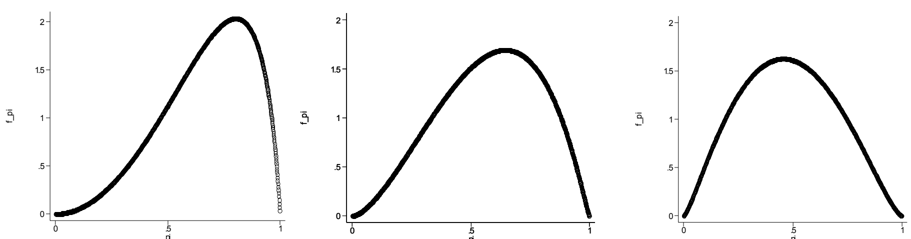
Imprecise Beliefs	.Average Beliefs	.Precise Beliefs
(focal index 1 std. dev. above mean)	(focal index at sample mean)	(focal index 1 std. dev. below mean)



Mode (Response): 0.71	Mode (Response): 0.65	Mode (Response): 0.62
Mean: 0.57	Mean: 0.58	Mean: 0.57
Median: 0.59	Median: 0.59	Median: 0.59
Variance of delta: 0.76	Variance of delta: 0.62	Variance of delta: 0.50

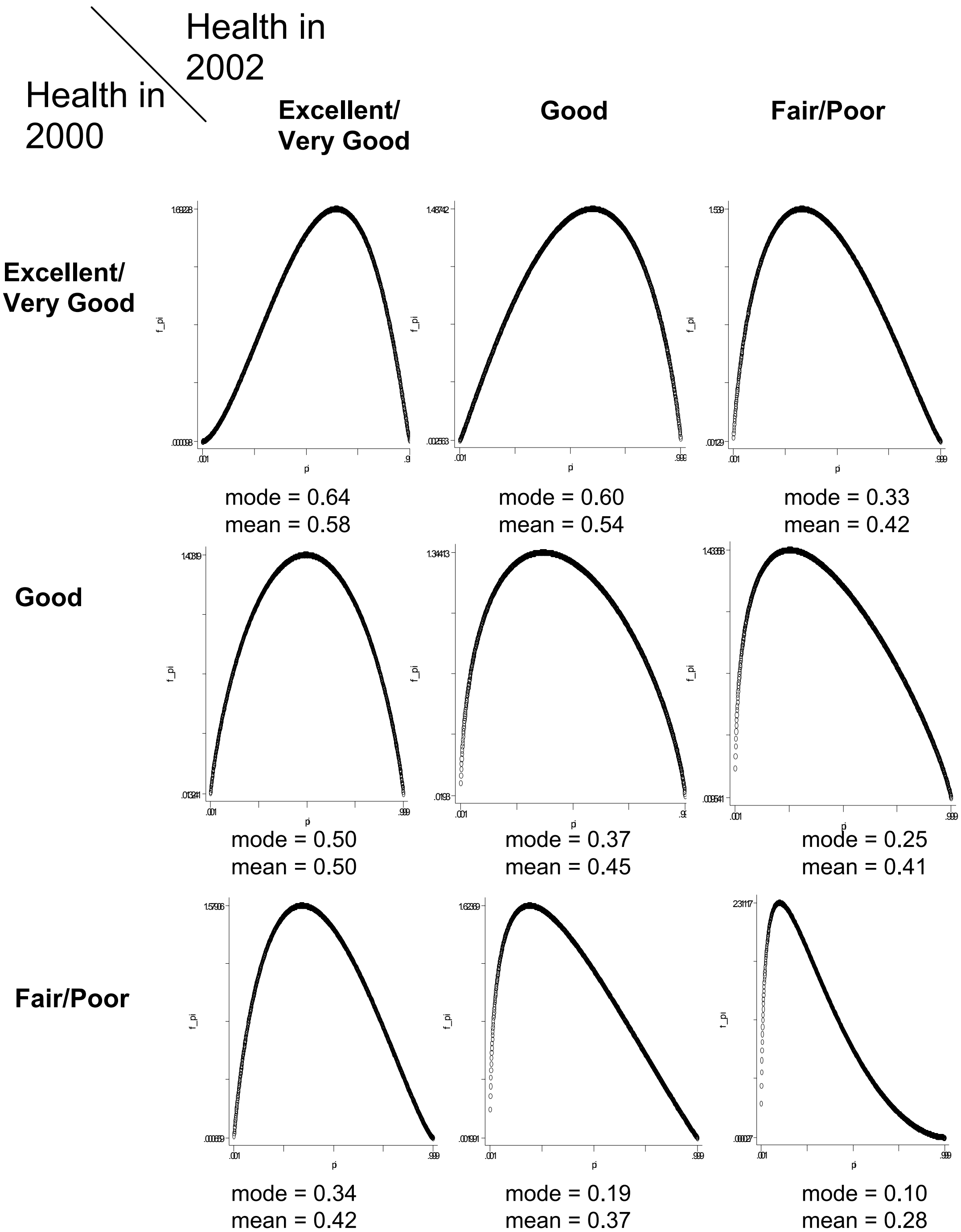
FIGURE 6: THE EFFECT OF A CHANGE IN OPTIMISM

Optimist's Beliefs	.Average Beliefs	.Pessimist's Beliefs
(optimism index 1 std. dev. above mean)	(optimism index at sample mean)	(optimism index 1 std. dev. below mean)



Mode (Response): 0.80	Mode (Response): 0.65	Mode (Response): 0.46
Mean: 0.67	Mean: 0.58	Mean: 0.48
Median: 0.70	Median: 0.59	Median: 0.47
Variance of delta: 0.62	Variance of delta: 0.62	Variance of delta: 0.62

Figure 7. Effect of Self-Reported Health and Changes in Health on Subjective Beliefs



Appendix I: Mathematical Derivation of the Likelihood Function

For convenience, the equations in the body of the paper are:

$$p = \Pr(I > 0 \mid \mu_i, \delta) = \Pr(\mu_i + \delta > u) = \Phi(\mu_i + \delta) \tag{1}$$

$$I = \mu_i + \delta - u \tag{2}$$

$$p = \Pr(\mu_i > u) = \Phi(\mu_i) \tag{3}$$

$$G(p) = \Pr(P < p) = \Pr(\Phi(\mu_i + \delta) < p) \tag{4}$$

$$= \Phi\left(\frac{\Phi^{-1}(p) - \mu_i}{\sigma_{\delta_i}}\right) \tag{5}$$

$$g(p) = \frac{\phi\left(\frac{\Phi^{-1}(p) - \mu_i}{\sigma_{\delta_i}}\right)}{\phi(\Phi^{-1}(p)) \cdot \sigma_{\delta_i}} \tag{6}$$

$$p = \Phi\left(\frac{\mu_i}{1 - \sigma_{\delta_i}^2}\right) \tag{7}$$

$$\mu = Z_p^* \beta_p^* \tag{8}$$

$$\ln \sigma_\delta = Z_a^* \beta_a^* \tag{9}$$

$$\mu = Z_p \beta_p + u_p \tag{10}$$

$$\ln \sigma_\delta = Z_a \beta_a + u_a \tag{11}$$

The Four Regions of the Graph

In order to use *MRH* to analyze a set of beliefs governed by equation (6), we calculate the value of p at which the derivative of $g(p)$ is zero. For $g(p)$ concave, this is the mode; for $g(p)$ convex, this is the minimum. The value of p at which this is true is

$$p = \Phi\left(\frac{Z_p^* \beta_p^*}{1 - \sigma_\delta^2}\right), \tag{A1.1}$$

where equation (8) has been used to substitute for μ.

According to *MRH*, our graph is divided into four contiguous regions based on the response coming from a respondent whose characteristics place her in that region. These four regions are the exact or non-heaped region, the zero region, the 100 region and the 50-50 region. These regions are shown in figure 3.

All possible probability densities governed by equation (6) are either uniform, strictly convex (u-shaped) or strictly concave (upside-down u-shaped), depending on the value of σ_{δ}. Given that, from equation (A1.1) it is clear that the convex densities reach a minimum and the concave densities reach a maximum in the interval $p \in (0,1)$. For clarity, in the main text of the paper we do not emphasize that the densities in the zero region or the 100 region have these properties. This is because we treat them as though the mode were at zero or 100 respectively, as described below.

Any respondent whose distribution is upside-down u-shaped and has a mode between 0.5% and 99.5% will give the mode, rounded to the nearest 1%, as her response. All such respondents fall into the exact region.

A respondent may fall into the zero region in one of two ways. First, if her distribution is upside-down u-shaped and the mode is less than 0.5%, then she will give a response of zero—respondents of this type are treated as though they give an exact answer of zero. Second, if her distribution is u-shaped and the minimum of *g(p)* occurs in the interval greater than 99.5%, then we assume that there is not enough probability weight on the right side (the side close to 100%) for the respondent to take it into account, and she will respond based on the left-side tail, hence giving a response of zero. The regions that cover these two sets of respondents are contiguous and hence the zero region is one continuous region of the graph.

Similarly, a respondent may fall into the 100 region in one of two ways. . First, if her distribution is upside-down u-shaped and the mode is greater than 99.5%, then she will give a response of 100%. Second, if her distribution is u-shaped and the minimum of *g(p)* occurs in the interval smaller than 0.5%, then we assume that there is not enough probability weight on the left side (the side close to 0) for the respondent to take it into account, and therefore she will respond based on the right-side tail, giving a response of 100%.

Finally, the 50-50 region covers all respondents who do not fall into the other regions. Distributions in this region are u-shaped, and the derivative of *g(p)* equals zero in the interval $p \in [0.5\%, 99.5\%]$. We assume, then, that neither tail dominates the other in the respondent's mind and that this leads to a response of 50%.

The Likelihood of an Exact Response

As can be seen by equations (6) and (8), a respondent's personal beliefs about a probability are completely determined by the pair $(Z^*_p\beta^*_p, \ln\sigma_\delta)$. Conversely, any possible modal response corresponds to a set of possible pairs $(Z^*_p\beta^*_p, \ln\sigma_\delta)$. For any set of observable factors and parameter values, $Z_p\beta_p, Z_a\beta_a$, the random variables u_a and u_p will induce a likelihood that the respondent's pair $(Z^*_p\beta^*_p, \ln\sigma_\delta)$ falls into any particular region of the graph.

In order to calculate a non-zero likelihood for an exact response, say 65%, it is necessary that we assume that such a response actually indicates that the respondent means that the probability is in the range (64.5%, 65.5%)—otherwise we face integrating a bivariate normal distribution over a single dimensional curve defined by the mode function in equation (A1.1). Using this assumption, we then solve for the likelihood of the set of pairs $(Z_p^* \beta_p^*, \ln \sigma_\delta)$ that yield distributions with modes in the relevant range.

The likelihood of a particular, exact-region response is:

$$L = \int_{-\infty}^{-Z_a \beta_a} \int_{B_1}^{B_2} f(u_p, u_a; \rho) du_p du_a \tag{A1.2}$$

Where *f* is the bivariate normal distribution of the two random variables, with unknown covariance ρ. Where

$$B_1 = \left(1 - e^{2(Z_a \beta_a + u_a)}\right) \cdot \Phi^{-1}(R - 0.5\%) - Z_p \beta_p$$

and

$$B_2 = \left(1 - e^{2(Z_a \beta_a + u_a)}\right) \cdot \Phi^{-1}(R + 0.5\%) - Z_p \beta_p .$$

In this example, R=65%. This region can be seen in Figure 3 as region A.

The Likelihood of a Focal Response

The likelihood of responses in the three heaped regions are simpler to calculate. In each region, it suffices to integrate the bivariate normal distribution of u_a and u_p over the entire region of interest. The boundaries of the regions are determined by the functions *h* and *k* as defined implicitly by equation (A1.1) and modal response hypothesis. These functions are shown in figure 3. Their equations are

$$k(Z_p^* \beta_p^*) = \frac{1}{2} \ln \left(1 - \frac{Z_p^* \beta_p^*}{\Phi^{-1}(0.995)} \right) \tag{A1.3}$$

and

$$h(Z_p^* \beta_p^*) = \frac{1}{2} \ln\left(1 - \frac{Z_p^* \beta_p^*}{\Phi^{-1}(0.005)}\right) \qquad \text{(A1.4)}$$

To calculate the likelihood of a zero response we see that the zero region extends from $-\infty$ to zero along the $\ln \sigma_\delta$-axis and from *h* to *k* along the $Z_p^* \beta_p^*$-axis. The likelihood of a zero response, then, is:

$$L = \int_{-\infty}^{-Z_p\beta_p} \int_{h-Z_a\beta_a}^{k-Z_a\beta_a} f(u_a, u_p; \rho) du_a du_p \qquad \text{(A1.5)}$$

In a similar fashion we see that the likelihood of a 100 response is

$$L = \int_{-Z_p\beta_p}^{\infty} \int_{k-Z_a\beta_a}^{h-Z_a\beta_a} f(u_a, u_p; \rho) du_a du_p \qquad \text{(A1.6)}$$

and the likelihood of a 50-50 response is

$$L = \int_{-\infty}^{-Z_p\beta_p} \int_{k-Z_a\beta_a}^{\infty} f(u_a, u_p; \rho) du_a du_p + \int_{-Z_p\beta_p}^{\infty} \int_{h-Z_a\beta_a}^{\infty} f(u_a, u_p; \rho) du_a du_p \qquad \text{(A1.7)}$$

Our complete likelihood function is the sum of the individual likelihoods for all respondents, using equations (A1.2), (A1.5), (A1.6) or (A1.7), depending on the response.

Appendix 2. Derivation of the Gompertz Survival Form

The "Gompertz" survival function from current age a to future time T that we have been working with is:

$$S(a,T) = e^{-\int_a^T m(x)dx} \qquad \text{(A2.1)}$$

where m(x) is the hazard of death at age x

$$m(x) = \mu e^{\gamma x} \qquad \text{(A2.2)}$$

where γ, and in general μ are allowed to vary across individuals—a feature which we will temporally ignore by setting $\mu = 1.0$. We integrate to obtain:

$$\int_a^T e^{\gamma x} dx = \frac{e^{\gamma T}}{\gamma} - \frac{e^{\gamma a}}{\gamma} \qquad \text{(A2.3)}$$

which after rearranging terms yields

$$\frac{e^{\gamma a}}{\gamma}\left[e^{\gamma(T-a)} - 1\right] \qquad \text{(A2.4)}$$

So:

$$S(a,T) = \exp\left[-\frac{e^{\gamma a}}{\gamma}\left[e^{\gamma(T-a)} - 1\right]\right] \qquad \text{(A2.5)}$$

Note that at *a=T, S(a,T)* = 1.0, while with $\gamma > 0$, *S(a,T)* approaches 0 as T approaches ∞.

To allow for imperfect discounting we add a parameter δ such that

$$S(a,T) = \exp\left[-\frac{e^{\gamma a}}{\gamma}\left[e^{\delta\gamma(T-a)} - 1\right]\right]. \qquad \text{(A2.6)}$$

To allow *S* to vary across individuals of the same age with the same horizon *(T-a)* We incorporate it into the modal response model by specifying the mode to be:

$$P_{\text{mode}} = \Phi\left(\frac{Z_p\beta_p - \Phi^{-1}(S(a,T))}{1-\sigma_\delta}\right) \qquad \text{(A2.7)}$$

So that the $\beta's$ are measures of the departure from the overall expected survival associated with the various covariates comprising Z_p. In spirit this is equivalent to allowing μ to vary systematically (with the covariates) across individuals.