Technical Appendix To:

Subjective Health Expectations: Correcting for Focal Point Bias

Kim P. Huynh[*] Juergen Jung[†]

Indiana University Towson University

21st April 2009

Abstract

We derive subjective health expectations using the RAND-HRS data. These expectations can be used in the estimation of structural life-cycle models. We use a Bayesian updating mechanism in order to correct for focal point responses and reporting errors of the orignial health expectations variable. In addition, we test the quality of the health expectations measure and describe its correlation with various health indicators and other individual characteristics. We find that subjective health expectations do contain additional information that is not incorporated in subjective mortality expectations and that the rational expectations assumption cannot be rejected for subjective health expectations.

JEL Classification: I10, D84, C11, C23

Keywords: Subjective Health Expectations, Rational Health Expectations, Work Limiting Health Problems, Bayesian Updating of Expectations.

[*]We would like to thank Gerhard Glomm, Michael Kaganovich, Wayne R. Gayle, Rusty Tchernis and Pravin Trivedi for many helpful comments. We are also grateful to Li Gan for making Matlab code available to us.

[†]Corresponding Author: Juergen Jung, Department of Economics, Stephensen Hall, Towson University, 8000 York Road, Towson, MD 21252-0001, phone: (812) 345-9182, e-mail: jjung@towson.edu

1 Appendix A: Descriptive Statistics and Construction of Health Tables using Population Health Hazard Rates

In table 9 we report summary statistics according to health status. The first panel in the table reports the proportions of individuals having a specific health status in wave 1 and wave 2. We see that 54.6% of people with excellent health in wave 1, do still report excellent health for wave 2, whereas 33.4% report their health status as very good and 0.2% report a decline in their health to the status of poor. Similarly, of the people with very good health in wave 1, 54.4% still have very good health in wave 2. In addition, 16.3% of those with very good health in wave 1 improved their health to the status of excellent in wave 2, whereas 25.1% saw their health decline to status "good". We see that health states are very persistent in the sense that for almost all health states 50% of the individuals remain in that stage.

Panel two in table 9 summarizes the mean expectations about work limiting health problems by health status. We find that individuals with better health status in both waves have lower expectations about future health problems. Individuals who could improve their health over the waves report lower subjective probabilities of future health problems. See panel 3 and the negative numbers in the upper right corner, where changes in expectations about future health are negative. Panel 4 and Panel 5 report the mean expectations of living to age 75 and age 85 respectively. We again see that individuals with a better health status report higher probabilities of surviving up to a target age.

When comparing wave 1 and wave 6 expectations according to health status we find that the persistence of health status over six waves is still quite strong. Although fewer individuals can maintain a health status of excellent over all six waves. We also find that people with the same health status in wave 6 have higher expectations to live to target age 75. This is what one would expect, given that these individuals are much older now, some of them probably very close to target age 75.

1.1 Procedure

Table 8 lists the percentage of those respondents who gave continuous responses, focal responses, and no responses in the first two waves. The table also reports transition probabilities of the different response modes over the first two waves. We see that in wave 1 only 41.76% of respondents gave continuous responses with 12.24% providing focal point responses. A relatively large section of respondents gave no answer to the expectations health question, 46.13%.

The focal point responses cannot represent respondents' true probabilities, so that without correcting for focal responses of zero or one, it is impossible to derive health curves that change over time. In this section we attempt to recover the "true" subjective health expectations curve for each respondent. We call these the adjusted subjective health expectations (curves).

We first derive health tables for the U.S. using observed outcome probabilities from the data. Manski (1993) has already suggested that outcome probabilities can be used as proxies for subjective mortality expectations. We then update these tables using the subjective health expectations. The resulting adjusted subjective health expectations do not contain focal point responses anymore but contain the additional information carried in the observed outcome probabilities (health tables).

In order to construct the health tables we first define the hazard rates for having a work limiting health problem as

$$\lambda_0(t) = \Pr(T = t_j | T \geq t_j) = \frac{d(t)}{l(t)}, \tag{1}$$

where $d(t)$ is the number of individuals developing a work limiting health problem at age t and $l(t)$ is the total number of individuals aged t without a health problem at the beginning of the period. The number of individuals developing a work limiting health problem from age t to $t+1$ is

$$d(t) = -[l(t+1) - l(t)].$$

A period in this context is the two year interval between waves in the Rand-HRS survey. The zero subscript in (1) denotes that the variable is derived from population realizations and not from a specific individual.

In addition we can derive the "survival probability". Survival in this context means remaining without a work limiting health problem from one period to the next. We denote this survival, or better, health maintenance probability from birth, as

$$S_0(t) = \Pr[T \geq t] = \prod\nolimits_{j|t_j \leq t} (1 - \lambda_j) = \frac{l(t)}{l(0)},$$

where $l(t)$ is again the number of individuals aged t without work limiting health problems and $l(0)$ is the starting cohort of newly bournes.

The health table "survival probability" from age a up to t without censoring is

$$S_{0a}(t) = \frac{S_0(a+t)}{S_0(a)} = \frac{\frac{l(a+t)}{l(0)}}{\frac{l(a)}{l(0)}} = \frac{l(a+t)}{l(a)}.$$

The health table hazard rate is the negative of the percentage change in the survival probability or more formally

$$\lambda_0(t) = -\Delta \ln S_0(t) = -\frac{1}{S_0(t)} \dot{S}_0(t) = -\frac{d \ln(S_0(t))}{dt} = -\%\Delta S_0(t).$$

We can also express this as

$$\lambda_0(t) = -\frac{S_0(t+1) - S_0(t)}{S_0(t)} = -\frac{\frac{l(t+1)}{l(0)} - \frac{l(t)}{l(0)}}{\frac{l(t)}{l(0)}} = -\frac{l(t+1) - l(t)}{l(t)} = \frac{d(t)}{l(t)}. \tag{2}$$

The cumulative health-problem hazard function (in continuous time) is[1]

$$\Lambda_0(t) = \int_0^t \lambda_0(t)\, d\tau = \int_0^t -\frac{d \ln(S_0(t))}{dt} d\tau = -\ln S_0(t). \tag{3}$$

[1] Compare also Venables and Ripley (2002) for formal details on hazard functions.

In figure 1 we report the health-hazard rates for men and women. We limit the sample to people who are 40 years of age and older. By assumption individuals start being at risk of a work limiting health problem at age 40. We then construct the Kaplan-Meier survival rate with 99% confidence bounds. We assume individuals live in good health (without work limiting health problems) until failure. Failure is defined as the onset of a work limiting health problem, given that no such prior condition existed. An individual who enters the survey with a health problem is assumed to have failed at the age of survey entry. An individual who recovers from a health problem and develops another health problem while still in the survey at a later age is counted again as having failed for that particular age group. An individual leaving the survey is a censored spell and decreases the number of individuals at risk without counting towards the number of failures.[2]

E.g. a 70 year old male entering the survey without a health problem and reporting a health problem at age $74, 76, 78$ is counted as having failed at age 74. If the same individual does not report a health problem at age 80, but again reports a problem at age 82, then a second failure is counted for the 82 year old age group. Similarly, if a 64 year old female enters the survey with a health problem, she is assumed to have failed at age 64.

We then count the number of people at risk at each age $l(t)$ where $t = 40, ..., 95$. Individuals at risk are all individuals in the survey that have not yet left the survey and do not have a health problem. In this sense, individuals who recover from a health problem but are still in the survey, will reenter the set of people at risk. We then count the number of people who fail at each age t, that is people who report a health limiting work problem at at t. The hazard rate for age t to $t+2$ is then defined as

$$\lambda(t) = \frac{d(t)}{l(t)} \equiv \lambda(t).$$

Since the hazard rates are very volatile we fit a 5^{th} order polynomial with least squares to smooth out the edges. From the top panel in figure 1 we see that the health hazard rates for men are higher than those for women over almost the entire age range. We will later report estimation results based on the original hazard rates and on the smoothed versions. We find that the results are robust and do not depend on whether we smooth the hazard functions before applying the Bayesian updating procedure. In figure **??** we also report unconditional hazard rates that we have calculated assuming that a person with a work limiting health problem in consecutive years is counted as having failed multiple times. The previous hazard rates would only count a transition from a healthy state to a sick state as failure which would then be re□ected in the hazard rate. If we count both transitions from healthy to sick and from sick to sick as failure then the resulting hazard rate will increase as we can see in figure **??**.

1.2 Subjective Hazard Rates and Survival Functions

We next turn our attention to the individual. The personal health-survival probability from age a to target age $a+t$ for individual i is $S_{ia}(t)$. Variable $S_{ia}(t)$ is a random variable and s_{iat} is a realization of this variable.[3]

[2]See (Cleves, Gould and Gutierrez, 2004, p. 59-62) for a discussion of how to model repeated failures by the same individual in Stata's survival package. Compare also (Cameron and Trivedi, 2005, p. 580 - 584) for a brief introduction to non-parametric survival analysis.

[3]We closely follow Gan, Hurd and McFadden (2003) and adopt their notation.

The density of random variable $S_{ia}(t)$ is $\pi(s_{ia}(t))$ or $\pi(s_{iat})$. The personal health-problem hazard rate at age a is denoted $\lambda_{ia}(t)$ and the cumulative hazard rate is $\Lambda_{ia}(t)$.

From (3) we can derive an individual $i's$ health "survival" probability (or health curve) as

$$S_{ia}(t) = \exp\left(-\Lambda_{ia}(a+t) + \Lambda_{ia}(a)\right) = \exp\left(-\int_0^t \lambda_{ia}(a+r)\,dr\right). \tag{4}$$

We next use an individual's response to the health related question in the interview asking for a probability of having a work limiting health problem within the next ten years. We denote this probability as $1 - p_{ia\tau}$, where i denotes the individual, a is the individual's age and τ is time. Then the survival probability, that is the probability of maintaining the good health status is $p_{ia\tau}$ and its density is conditional on the personal survival probability from age a to age $a + t$ as in

$$f\left(p_{ia\tau} | S_{ia\tau} = s_{ia\tau}\right).$$

The method employed uses the population hazard function $\lambda_{0a}(a+t)$ as a base and modifies it to calculate individual hazard rates $\lambda_{ia}(a+t)$ according to the following hazard scaling function

$$\lambda_{ia}(a+t) = \gamma_i \lambda_{0a}(a+t), \tag{5}$$

where $\gamma_i > 1$ indicates a "pessimistic" and a $\gamma_i < 1$ an "optimistic" individual.[4]

With focal responses and response errors present in $p_{ia\tau}$ the personal survival curve is not forced through $p_{ia\tau}$ at $a + \tau$. In this case we employ a Bayesian approach to update the individual survival curve. We denote the prior belief about the personal survival curve density as $\pi(s_{iat})$. The mean of the prior density is $\exp(-\Psi\Delta\Lambda_{0at})$ and its standard deviation is σ_2. Parameter Ψ measures the population's average optimistic degree. Given S_{iat}, the self-reported survival probability p_{iat} has density $f(p_{iat}|s_{iat})$ so that the difference between the survival probability S_{iat} and the self-reported survival probability p_{iat} is the measurement error. We use the observed $p_{ia\tau}$ to update the prior density $\pi(s_{ia\tau})$ in order to obtain the posterior density $\pi(s_{ia\tau}|p_{ia\tau})$. The posterior density is given by

$$\pi\left(s_{ia\tau}|p_{ia\tau}\right) = \frac{f\left(p_{ia\tau}|s_{ia\tau}\right)\pi\left(s_{ia\tau}\right)}{\int f\left(p_{ia\tau}|s_{ia\tau}\right)\pi\left(s_{ia\tau}\right)ds_{ia\tau}},$$

with mean μ_{ia} and standard deviation σ_1. It can be shown that the best estimator for $S_{i\tau}$ with a quadratic loss function $L\left(S_{it}, \hat{S}_{it}\right) = E\left[S_{it} - \hat{S}_{it}\right]^2$ is the conditional expectation, so that

$$\hat{S}_{i\tau} = E\left(S_{i\tau}|p_{ia\tau}\right).$$

We then apply $\hat{S}_{i\tau}$ to the observed record of realized health problems to obtain the model's parameters σ_1, σ_2

[4] Gan, Hurd and McFadden (2003) also calculate an age scaling model which leads to inferior results. We therefore concentrate on the hazard scaling version of their model.

and Ψ. The log-likelihood function is given as

$$\ln L = \sum_{\text{NoHealthProblems}} \ln \hat{S}_{it} + \sum_{\text{HealthProblems}} \ln\left(1 - \hat{S}_{it}\right). \tag{6}$$

We next make some assumption concerning the prior distribution of random variable S_{iat}. We denote the distribution of S_{iat} as $\pi\left(s_{ia\tau}\right)$ and define it as a truncated normal distribution. The mean of S_{iat} is $\exp\left(-\Psi\Delta\Lambda_{0at}\right)$, the variance is σ_2^2 and the truncation is between $0 < s_{ia} < 1$. The prior distribution is

$$\pi\left(s_{ia};\Psi\right) = \frac{\frac{1}{\sigma_2}\phi\left(\frac{s_{ia}-v_{ia}}{\sigma_2}\right)}{\Phi\left(\frac{1-v_{ia}}{\sigma_2}\right) - \Phi\left(-\frac{v_{ia}}{\sigma_2}\right)},$$

where v_{ia} is the mean and σ_2 the standard deviation of the normal distribution. Both values satisfy

$$\exp\left(-\Psi\Delta\Lambda_{0a\tau}\right) = v_{iat} - \sigma_2\eta\left(0,1,v_{iat},\sigma_2\right).$$

The right hand side is the mean of the truncated normal according to the formula in the appendix.

The conditional density of the responses to interview survival questions is assumed to follow a censored normal distribution

$$\begin{aligned} f\left(p_{ia\tau}|s_{ia\tau}\right) &= \phi\left(\frac{p_{ia\tau}-\mu_{ia\tau}}{\sigma_1}\right) \text{ when } 0 < p_{ia\tau} < 1, \\ \Pr\left(p_{ia\tau}=0|s_{ia\tau}\right) &= 1-\Phi\left(\frac{\mu_{ia\tau}}{\sigma_1}\right), \text{ and} \\ \Pr\left(p_{ia\tau}=1|s_{ia\tau}\right) &= 1-\Phi\left(\frac{1-\mu_{ia\tau}}{\sigma_1}\right), \end{aligned}$$

with variance σ_1^2. The expected value $E\left[S_{ia}\right]$ of the conditional distribution is

$$s_{ia} = 0 \times \Pr\left(p_{ia\tau}=0|s_{ia\tau}\right) + E\left[x|0<x<1\right] \times f\left(p_{ia\tau}|s_{ia\tau}\right) + 1 \times \Pr\left(p_{ia\tau}=1|s_{ia\tau}\right),$$

so that

$$s_{ia} = \left[\Phi\left(\frac{1-\mu_{ia}}{\sigma_1}\right) + \Phi\left(\frac{\mu_{ia}}{\sigma_1}\right) - 1\right]\left[\mu_{ia} - \sigma_1\eta\left(0,1,\mu_{ia},\sigma_1\right)\right] + \left[1-\Phi\left(\frac{1-\mu_{ia}}{\sigma_1}\right)\right],$$

where it can be shown (see Appendix A) that $E\left[x|0<x<1\right] = \left[\mu_{ia} - \sigma\eta\left(0,1,\mu_{ia},\sigma_1\right)\right]$. Finally, given $p_{ia\tau}$, the posterior distribution is given by

$$\pi\left(s_{ia}|p_{ia\tau}\right) = \frac{f\left(p_{ia\tau}|s_{ia\tau}\right)\pi\left(s_{ia\tau}\right)}{\int f\left(p_{ia\tau}|s_{ia\tau}\right)\pi\left(s_{ia\tau}\right)ds_{ia\tau}}.$$

Then the best estimator for S_{ia} under a mean square loss function is its mean, that is

$$\hat{S}_{ia} = E\left[S_{ia}\right] = \int_0^1 s_{ia}\pi\left(s_{ia}|p_{ia\tau}\right)ds_{ia} = \frac{\int_0^1 s_{ia}\phi\left(\frac{p_{ia\tau}-\mu_{ia\tau}(s_{ia},\sigma_1)}{\sigma_1}\right)\phi\left(\frac{s_{ia}-v_{ia}(\Psi,\sigma_2)}{\sigma_2}\right)ds_{ia}}{\int \phi\left(\frac{p_{ia\tau}-\mu_{ia\tau}(s_{ia},\sigma_1)}{\sigma_1}\right)\phi\left(\frac{s_{ia}-v_{ia}(\Psi,\sigma_2)}{\sigma_2}\right)ds_{ia\tau}}.$$

We get similar results for the focal point responses at $p_{iat} = 0$ and 1 so that we summarize the predicted survival probabilities as

$$\hat{S}_{ia} = \begin{cases} \frac{\int_0^1 s_{ia}\left(1-\Phi\left(\frac{\mu_{ia}(s_{ia},\sigma_1)}{\sigma_1}\right)\phi\left(\frac{s_{ia}-v_{ia}(\Psi,\sigma_2)}{\sigma_2}\right)\right)ds_{ia}}{\int_0^1\left(1-\Phi\left(\frac{\mu_{ia}(s_{ia},\sigma_1)}{\sigma_1}\right)\phi\left(\frac{s_{ia}-v_{ia}(\Psi,\sigma_2)}{\sigma_2}\right)\right)ds_{ia}}, \text{ if } p_{iat}=0 \\ \\ \frac{\int_0^1 s_{ia}\phi\left(\frac{p_{ia\tau}-\mu_{ia\tau}(s_{ia},\sigma_1)}{\sigma_1}\right)\phi\left(\frac{s_{ia}-v_{ia}(\Psi,\sigma_2)}{\sigma_2}\right)ds_{ia}}{\int \phi\left(\frac{p_{ia\tau}-\mu_{ia\tau}(s_{ia},\sigma_1)}{\sigma_1}\right)\phi\left(\frac{s_{ia}-v_{ia}(\Psi,\sigma_2)}{\sigma_2}\right)ds_{ia\tau}}, \text{ if } 0 < p_{iat} < 1 \\ \\ \frac{\int_0^1 s_{ia}\left(1-\Phi\left(\frac{1-\mu_{ia}(s_{ia},\sigma_1)}{\sigma_1}\right)\phi\left(\frac{s_{ia}-v_{ia}(\Psi,\sigma_2)}{\sigma_2}\right)\right)ds_{ia}}{\int_0^1\left(1-\Phi\left(\frac{1-\mu_{ia}(s_{ia},\sigma_1)}{\sigma_1}\right)\phi\left(\frac{s_{ia}-v_{ia}(\Psi,\sigma_2)}{\sigma_2}\right)\right)ds_{ia}}, \text{ if } p_{iat}=1. \end{cases} \tag{7}$$

Since respondents are interviewed every two years we can update the predictions according to whether they are still without work limiting health problems. Then the likelihood function changes from (6) to

$$\ln L = \sum_{\text{NoHealthProblems}} \ln \hat{S}_{ia2} + \sum_{\text{HealthProblems}} \ln\left(1-\hat{S}_{ia2}\right). \tag{8}$$

From (4) and (5) one can calculate the optimism parameter γ_i as

$$\begin{aligned} \hat{S}_{ia}(t) &= \exp\left(-\int_0^t \hat{\gamma}_i\lambda_{0a}(a+r)\,dr\right), \\ &\rightarrow \hat{S}_{ia}(t) = \exp\left(-\hat{\gamma}_i\Delta\Lambda_{0a}(t)\right). \end{aligned}$$

Taking logs we can solve for $\hat{\gamma}_i$ as

$$\hat{\gamma}_i = -\frac{\ln \hat{S}_{ia\tau}}{\Delta\Lambda_{0a\tau}},$$

so that

$$\hat{S}_{ia2} = \hat{S}_{ia\tau}^{\left(\frac{\Delta\Lambda_{0a2}}{\Delta\Lambda_{0a\tau}}\right)}. \tag{9}$$

Substituting (9) into the log-likelihood function (8) we have

$$\ln L = \sum_{\text{NoHealth Pr oblems}} \ln \hat{S}_{ia\tau}^{\left(\frac{\Delta\Lambda_{0a2}}{\Delta\Lambda_{0a\tau}}\right)} + \sum_{\text{Health Pr oblems}} \ln\left(1-\hat{S}_{ia\tau}^{\left(\frac{\Delta\Lambda_{0a2}}{\Delta\Lambda_{0a\tau}}\right)}\right). \tag{10}$$

For further details on these derivations we refer to Gan, Hurd and McFadden (2003).

1.3 Estimation Results

We use a subset of the data to estimate the likelihood function in expression 10. We only use wave 1 and wave 2 in order to contain the computation burden. We only keep observations where respondents report no work limiting health problem in wave 1. This reduces the data to 7001 observations, 3489 of which are males and 3512 of which are females.

We report estimation results for two separate models in table 2. The first is a restricted model where we set $\Psi = 1$ and estimate σ_1 and σ_2. In this case the mean of the prior distribution is equal to the realizations in the health-tables. We report standard errors in parenthesis. Standard errors where obtained using a Bootstrap routine on 500 subsamples with 400 observations each. The first column uses Health Table data using a 5th degree polynomial to smooth the Kaplan-Meier estimate of the survival curve. The second column uses the original Kaplan-Meier estimator for the health table survival curve. Finally, in column three we report the estimation results for the unrestricted model where parameter Ψ is also estimated. We find that $\hat{\Psi} = 2.37$ which indicates that individuals are much more pessimistic about their health than the objective realization rates in the health tables.

Finally, we construct the health curves using the estimates of the restricted model. The top panel of figure 3 displays the health survival probabilities (the probability of remaining without work limiting health problems) for a 50 year old man. The blue line depicts the survival rates of an individual claiming a 100 percent change of staying in good health (or a 0 percent chance of developing a work limiting health problem), whereas the red line is an individual stating a 0 chance of staying in good health within the next 10 years. The green line is the subjective survival rate of an individual with average expectations about her health. The solid black line is the health-table survival rate. Figure 5 displays the analog results for 60 year old individuals.

In addition, we plot the confidence bounds of the health table estimates. We see that the confidence bounds of the adjusted subjective health curves of individuals reporting $p_{iat} = 0$ or 1 lie well beyond the confidence bounds of the health table estimates. Therefore, a model using the health table realizations as proxies for subjective expectations neglects statistically significant information from subjective expectations.

Figures 4 and 6 plot the survival curves for the unrestricted model. We see that in this model agent are more pessimistic, which is re□ected in the estimate of $\hat{\Psi} = 2.37$ and the lower subjective survival curves. We report the histogram of self reported health expectations, together with the histogram of self reported health expectations after adjusting for focal point responses using the restricted model and the unrestricted model in figure **??**. We see that the focal point responses at 0 and 1 have disappeared and that the unrestricted model exhibits the more pessimistic subjective health expectations.

1.4 Algorithm

We would like to thank Li Gan for making Matlab code available to us. We next describe our implementation of the algorithm. This implementation differs from Gan's code in the sense that we needed to construct the outcome probabilities (recorded in Health Tables) first. We also restrict my attention to the hazard scaling model.

1. Construct health tables using the population realizations of the hazard rate λ for each age group a of the form
$$\lambda_{0a}(a) = \frac{d(a)}{l(a)}.$$

2. Use individual data on subjective expectations about work limiting problems within the next 10 years, denoted as $ExpHealthProblems = (1 - p_{ia})$, so that the probability of NOT having a work limiting health problem is p_{ia}. We interpret this also as the perceived survival rate (survival in 'good health') of individual i at age a.

3. Create dummy variable $d_{i,a,a+2} = 1$ if individual i was in good health in period 1 at age a and is still in good health in period 2 at age $a + 2$ and $d_{i,a,a+2} = 0$ otherwise.

4. Calculate the cumulative hazard rate $\Lambda_{0a}(a + 10)$ up to the target age $a + 10$. The target age is $a + 10$ because p_{ia} is defined as the subjective belief about surviving 10 years without work limiting health problems. We use
$$\Lambda_{0a}(a + 10) = \sum\nolimits_{t=1}^{10} \lambda_{0a}(a + t).$$

5. Calculate the cumulative hazard rate $\Lambda_{0a}(a + 2)$ up to the next wave at age $a + 2$ which is
$$\Lambda_{0a}(a + 2) = \sum\nolimits_{t=1}^{2} \lambda_{0a}(a + t).$$

6. Likelihood Routine:

   (a) Solve for μ_{ia} out of
$$s_{ia} = \left[\Phi\left(\frac{1-\mu_{ia}}{\sigma_1}\right) + \Phi\left(\frac{\mu_{ia}}{\sigma_1}\right) - 1\right][\mu_{ia} - \sigma_1\eta(0, 1, \mu_{ia}, \sigma_1)] + \left[1 - \Phi\left(\frac{1-\mu_{ia}}{\sigma_1}\right)\right]. \quad (11)$$
   Where s_{ia} is a grid vector from $[0, ..., 1]$ and therefor μ_{ia} is also a vector.

   (b) Solve for v_{iat} out of
$$\exp(-\Psi\Lambda_{ia}(a + 10)) = v_{iat} - \sigma_2\eta(0, 1, v_{iat}, \sigma_2). \quad (12)$$

   (c) Solve for $\hat{S}_{iat}$ distinguishing $p_{iat} = 0, 1$, or interior from (7).

   (d) Build log-likelihood function from
$$\ln L(\sigma_1, \sigma_2, \Psi) = \sum_{i=1}^{N}\left[d_{i,a,a+2}\ln \hat{S}_{ia\tau}^{\left(\frac{\Lambda_{ia}(a+2)}{\Lambda_{ia}(a+10)}\right)} + (1 - d_{i,a,a+2})\ln\left(1 - \ln \hat{S}_{ia\tau}^{\left(\frac{\Lambda_{ia}(a+2)}{\Lambda_{ia}(a+10)}\right)}\right)\right].$$

   (e)
$$\left(\hat{\sigma}_1, \hat{\sigma}_2, \hat{\Psi}\right) = \arg\max_{\{\sigma_1,\sigma_2,\Psi\}} \ln L\left(\sigma_1, \sigma_2, \Psi|\hat{S}_{iat}\right).$$
   The restricted model fixes $\Psi = 1$ and only estimates σ_1 and σ_2.

7. Construction of subjective health curves:

   (a) Given $\left(\hat{\sigma}_1, \hat{\sigma}_2, \hat{\Psi}\right)$ solve for μ_{ia} and v_{iat} from (11) and (12).

   (b) Calculate estimates for survival $\hat{S}_{at}\left(p_{at}=0\right), \hat{S}_{at}\left(p_{at}=\bar{p}\right)$ and $\hat{S}\left(p_{at}=1\right)$ from (7), where $\bar{p}$ is the average subjective probability of surviving in good health of the sample.

   (c) Calculate the cumulative hazard rates from the hazard rates starting at a certain base age a so that

$$
\begin{aligned}
\Lambda_{0a}(a) &= \lambda_{0a}(a), \\
\Lambda_{0a}(a+1) &= \lambda_{0a}(a) + \lambda_{0a}(a+1), \\
&\vdots \\
\Lambda_{0a}(a+T) &= \sum\nolimits_{t=0}^{T} \lambda_{0a}(a+t).
\end{aligned}
$$

   Then define the following vector

$$
\Lambda_{0aT} = \left[\Lambda_{0a}(a), \Lambda_{0a}(a+1), ..., \Lambda_{0a}(a+T)\right].
$$

   So that the vector of survival rates in good health from age a to age $a+T$ is

$$
S_{0aT} = \exp\left(-\Lambda_{0aT} + \lambda_{0a}(a)\right).
$$

   The addition of the initial hazard rate normalizes the survival function S_{0aT} to be equal to 1 at age a. The zero subscripts denote the fact that these are mortality rates and survival rates of the population and not of a particular individual. We denote vector S_{0aT} to be the health table (population) survival rate of an individual with age a up to age $a+T$.

   (d) we finally update the health table survival rate with the subjective survival probability from the data $p_{ia\tau}$ using the hazard scaling model described earlier $\lambda_{ia}(a+t) = \gamma_i \lambda_{0a}(a+t)$. Where the estimate of γ for a particular individual i, aged a who answers with $p_{ia\tau}$ for the health expectations questions is

$$
\hat{\gamma}_i\left(p_{ia\tau}\right) = -\frac{\ln \hat{S}_{ia\tau}\left(p_{ia\tau}\right)}{\Lambda_{a0}(a+10)},
$$

   where $\hat{S}_{ia\tau}\left(p_{ia\tau}\right)$ was calculated in step (b) above.

   (e) The vector of subjective survival rates in good health is then

$$
S_{iaT}\left(p_{ia\tau}\right) = \exp\left(-\hat{\gamma}_i\left(p_{ia\tau}\right)\overbrace{\left[-\Lambda_{0aT} + \lambda_{0a}(a)\right]}^{S_{0aT}}\right),
$$

   where we plot these rates for $p_{ia\tau} = 0, 1$ and $\bar{p}$ in figure 3 for $a = 50$ and in figure 5 for $a = 60$.

1.5 Propositions[5]

Proposition 1 *(Mean of the truncated normal) If* $x\tilde{\ }N\left[\mu,\sigma^2\right]$ *and* e *and* f *are constant, then*

$$\begin{aligned} E\left[x|e\leq x\leq f\right] &= \mu-\sigma\eta\left(e,f,\mu,\sigma\right), \text{ where} \\ \eta\left(e,f,\mu,\sigma\right) &= \frac{\phi\left(\frac{f-u}{\sigma}\right)-\phi\left(\frac{e-u}{\sigma}\right)}{\Phi\left(\frac{f-u}{\sigma}\right)-\Phi\left(\frac{e-u}{\sigma}\right)}. \end{aligned}$$

Proposition 2 *(Mean of the censored normal) If* $x^*\tilde{\ }N\left[\mu,\sigma^2\right]$ *and*

$$x=\begin{cases} e & \text{if } x^*\leq e \\ x^* & \text{if } e\leq x^*\leq f \\ f & \text{if } f\leq x^* \end{cases},$$

where e *and* f *are constant, then*

$$E\left[x\right]=\Phi\left(\frac{e-\mu}{\sigma}\right)e+\left[\Phi\left(\frac{f-\mu}{\sigma}\right)-\Phi\left(\frac{e-\mu}{\sigma}\right)\right]\left[\mu-\sigma\eta\left(e,f,\mu,\sigma\right)\right]+\left[1-\Phi\left(\frac{f-\mu}{\sigma}\right)\right]f.$$

Proposition 3 *When* $p_{iat}=0$, *then*

$$\hat{S}_{ia}=\frac{\int_0^1 s_{ia}\left(1-\Phi\left(\frac{\mu_{ia}(s_{ia},\sigma_1)}{\sigma_1}\right)\phi\left(\frac{s_{ia}-v_{ia}(\Psi,\sigma_2)}{\sigma_2}\right)\right)ds_{ia}}{\int_0^1\left(1-\Phi\left(\frac{\mu_{ia}(s_{ia},\sigma_1)}{\sigma_1}\right)\phi\left(\frac{s_{ia}-v_{ia}(\Psi,\sigma_2)}{\sigma_2}\right)\right)ds_{ia}}.$$

Proposition 4 *When* $p_{iat}=1$, *then*

$$\hat{S}_{ia}=\frac{\int_0^1 s_{ia}\left(1-\Phi\left(\frac{1-\mu_{ia}(s_{ia},\sigma_1)}{\sigma_1}\right)\phi\left(\frac{s_{ia}-v_{ia}(\Psi,\sigma_2)}{\sigma_2}\right)\right)ds_{ia}}{\int_0^1\left(1-\Phi\left(\frac{1-\mu_{ia}(s_{ia},\sigma_1)}{\sigma_1}\right)\phi\left(\frac{s_{ia}-v_{ia}(\Psi,\sigma_2)}{\sigma_2}\right)\right)ds_{ia}}.$$

References

Cameron, Colin A. and Pravin K. Trivedi. 2005. *Microeconometrics, Methods and Applications*. New York: Cambridge University Press.

Cleves, Mario A., William W. Gould and Roberto G. Gutierrez. 2004. *An Introduction to Survival Analysis Using Stata*. Revised edition ed. Stata Press.

Gan, Li, Michael Hurd and Daniel McFadden. 2003. "Individual Subjective Survival Curves." NBER Working Paper 9480.

Manski, Charles F. 1993. "Dynamic Choice in Social Settings: Learnings from the Experiences of Others." *Journal of Econometrics* 58(2):121–136.

[5] We brie□y state the following propositions without proofs. Proofs can be found in Gan, Hurd and McFadden (2003).

Venables, William N. and B. Ripley. 2002. *Modern Applied Statistics with S*. 4 ed. New York: Springer Verlag.

2 Appendix B: Additional Tables

Wave	Year	Number of Obs.	%	Died	%
1	1992	12,652	9.31	229	1.8
2	1994	19,871	14.62	1,061	5.3
3	1996	19,052	14.02	1,224	6.4
4	1998	22,608	16.64	1,321	5.8
5	2000	20,900	15.38	1,411	6.8
6	2002	19,577	14.40	1,106	5.6
7	2004	21,245	15.63	–	–
Total	–	135,905	100.00	6,352	

Table 1: Observations by Wave and Number of Deceased

	1 Restricted	2 Restricted	3 Unrestricted
σ_1 :	0.0157 (0.0059)	0.0229 (0.0082)	0.0241 (0.0133)
σ_2 :	0.0453 (0.0124)	0.0397 (0.0091)	0.0191 (0.0091)
Ψ :	1	1	2.3706 (0.5119)
LogLikelihood:	−3956.821	−2953.354	−2881.642
Sample Size:	7001	7001	7001
Data Type:	5th deg. polyn	original	original
Iterations;	146	90	250
Estimation Time:	1.39 hours	0.81 hours	2.11 hours

Table 2: Estimation Results from Maximum Likelihood Estimation

Table 3: Smoker and Non-Smoker Health Expectations (Age Group: 40-60 in Wave 1)

	NumSmokers	Mean	StDev	NumOfNonsmokers	Mean	StDev
Wave1						
ExpHealthProblem	1642	40.749086	28.292658	4666	37.558937	27.533459
ExpLive to 75	1642	62.411693	30.296621	4666	68.671239	26.57337
ExpLive to 85	1642	40.310597	32.321715	4666	47.038148	30.684465
Wave2						
ExpHealthProblem	1494	36.92905	29.148447	4959	35.037911	27.058088
ExpLive to 75	1494	65.617805	26.106905	4959	69.51462	23.875841
ExpLive to 85	1494	41.838688	31.310986	4959	45.831216	29.373332
Wave3						
ExpHealthProblem	924	40.501082	28.986772	3455	38.356295	27.880392
ExpLive to 75	924	66.831169	26.224501	3455	71.804052	24.086389
ExpLive to 85	924	41.928571	32.143491	3455	49.053546	30.487931
Wave4						
ExpHealthProblem	714	41.239496	27.641363	2888	38.740651	27.088322
ExpLive to 75	714	65.12465	26.136762	2888	70.82964	24.022188
ExpLive to 85	714	39.752101	31.653611	2888	47.154778	29.569266
Wave5						
ExpHealthProblem	537	45.013035	28.446694	2381	42.102058	26.497837
ExpLive to 75	537	61.837989	28.732454	2381	70.662747	24.434359
Wave6						
ExpHealthProblem	368	44.758152	27.23845	1709	42.282036	27.143992
ExpLive to 75	368	63.179348	29.684727	1709	70.034523	25.442604

Table 4: Health Expectations per Income Quantiles (Age Group: 40-60 in Wave 1)

	1stQuantile	2ndQuantile	3rdQuantile	4thQuantile
Wave 1				
ExpHealthProblems				
Number of observations	353	1843	2320	2165
Mean	41.586402	41.220836	39.262931	35.307159
StDev	29.364075	28.887752	27.716295	26.313206
Wave 2				
ExpHealthProblems				
Number of observations	1330	1436	2067	2134
Mean	41.080451	37.771588	34.859216	32.917994
StDev	31.020011	29.023901	27.340296	25.876701
Wave 3				
ExpHealthProblems				
Number of observations	758	423	1799	1855
Mean	42.217678	41.193853	39.017232	37.32938
StDev	30.801171	29.67007	29.43682	26.927704
Wave 4				
ExpHealthProblems				
Number of observations	694		126	1697
Mean	40.992795		38.071429	40.721273
StDev	29.18397		28.057136	27.801002
Wave 5				
ExpHealthProblems				
Number of observations	621		1199	1590
Mean	42.832528		46.589658	40.909434
StDev	27.562029		27.391463	26.217142
Wave 6				
ExpHealthProblems				
Number of observations	595		748	1481
Mean	46.583193		46.533422	42.27684
StDev	28.072256		28.475833	27.644662
StDev				

Table 5: Percentage of Individuals according to Wave 1 and Wave 3 Expectations. (Age Group: 40-60 in Wave 1)

	r2>r3	r2<r3	r2=r3	Total	r2=r3=0	r2=r3=50	r2=r3=100
ExpHealthProblem	37.89	39	23.11	100	6.05	11.63	.76
ExpLive to 75	37.58	34.72	27.7	100	2.06	10.49	8.9
ExpLive to 85	42.35	38.64	19.02	100	3.11	6.23	3.43

Table 6: Percentage of Individuals according to Wave 1 and Wave 4 Expectations. (Age Group: 40-60 in Wave 1)

	r3>r4	r3<r4	r3=r4	Total	r3=r4=0	r3=r4=50	r3=r4=100
ExpHealthProblem	63.03	22.03	14.94	100	3.63	8.22	.45
ExpLive to 75	58.66	21.21	20.13	100	1.38	7.24	7.08
ExpLive to 85	63.88	22.42	13.71	100	2.03	4.44	2.56

Table 7: Health Status in Wave 1 (row) and Wave 6 (column): (1) health Transition probabilities, (2) Mean work limiting health expectations by health status in wave 1 and wave 2, (3) Mean change in work limiting health expectations by health status in wave 1 and wave 2, and (4) Mean mortality expectations to age 75. The column entries depict the health status in wave 6. (Age Group: 40-60 in Wave 1)

	1 Excellent	2 Very Good	3 Good	4 Fair	5 Poor
(1) Transition probabilites					
1 Excellent	.364	.111	.047	.014	.211
2 Very Good	.423	.477	.253	.079	.316
3 Good	.177	.338	.504	.36	.316
4 Fair	.03	.068	.183	.511	.158
5 Poor	.005	.006	.013	.036	
Total	1	1	1	1	1
(2) Mean health expectations					
1 Excellent	.229	.324	.393	.85	.4
2 Very Good	.287	.36	.389	.645	.817
3 Good	.289	.375	.425	.538	.65
4 Fair	.458	.432	.446	.521	.533
5 Poor	.375	.24	.513	.48	
(3) Mean change in health exp					
1 Excellent	.079	-.003	-.002	-.35	-.087
2 Very Good	.102	.074	.049	-.032	-.225
3 Good	.156	.129	.079	-.051	-.267
4 Fair	.102	.066	.118	.048	.3
5 Poor	.35	.55	.25	.08	
(4) Mean live75 expectations					
1 Excellent	82.935	72.308	74.839	87.5	15
2 Very Good	75.014	71.061	66.975	59.5	41.111
3 Good	74.933	69.142	58.899	52.537	25.833
4 Fair	73.704	67.813	63.167	46.977	50
5 Poor	80	61.429	44	36.25	

Table 8: Focal Responses and Continuous Responses about Work Limiting Health Expectations and the Transitions from Wave 1 to Wave 2

	continuous	0	100	NA	Total
Wave1					
Continuous	4078	655	141	479	5353
Continuous in percent	76.18	12.24	2.63	8.95	41.76
0	710	401	33	120	1264
0 in percent	56.17	31.72	2.61	9.49	9.86
100	167	35	48	38	288
100 in percent	57.99	12.15	16.67	13.19	2.25
NA	40	15	2	5856	5913
NA in percent	.68	.25	.03	99.04	46.13
Total	4995	1106	224	6493	12818
Total in percent	38.97	8.63	1.75	50.66	100

3 Appendix C: Additional Figures

Table 9: Health Status in Wave 1 (row) and Wave 2 (column): (1) health Transition probabilities, (2) Mean work limiting health expectations by health status in wave 1 and wave 2, (3) Mean change in work limiting health expectations by health status in wave 1 and wave 2, (4) Mean mortality expectations to age 75 and (5) Mean mortality expectations to age 85. The column entries depict the health status in wave 2. (Age Group: 40-60 in Wave 1)

	1 Excellent	2 Very Good	3 Good	4 Fair	5 Poor
(1) Transition probabilites					
1 Excellent	.546	.163	.05	.035	.014
2 Very Good	.334	.544	.262	.084	.083
3 Good	.102	.251	.544	.345	.097
4 Fair	.017	.039	.128	.471	.444
5 Poor	.002	.004	.016	.065	.361
Total	1	1	1	1	1
(2) Mean health expectations					
1 Excellent	.256	.359	.408	.3	.7
2 Very Good	.295	.364	.39	.514	.533
3 Good	.369	.378	.444	.54	.486
4 Fair	.415	.403	.482	.529	.759
5 Poor	.067	.586	.488	.564	.631
(3) Mean change in health exp					
1 Excellent	-.016	-.097	-.122	.037	-.7
2 Very Good	.009	-.049	-.04	-.03	-.317
3 Good	-.039	.011	-.051	-.101	.114
4 Fair	.011	.079	.007	-.031	-.106
5 Poor	.4	.236	.13	.161	.058
(4) Mean live75 expectations					
1 Excellent	79.314	70.427	71.053	39.524	70
2 Very Good	74.235	70.258	66.116	61.765	70
3 Good	71.324	67.202	61.115	52.895	36.923
4 Fair	61.667	64.773	58.017	49.431	36.471
5 Poor	90	76	54.595	46.078	41.311
(5) Mean live85 expectations					
1 Excellent	52.013	52.013	52.013	52.013	52.013
2 Very Good	50.441	50.441	50.441	50.441	50.441
3 Good	50.441	50.441	50.441	50.441	50.441
4 Fair	34.667	34.667	34.667	34.667	34.667
5 Poor	56.667	56.667	56.667	56.667	56.667

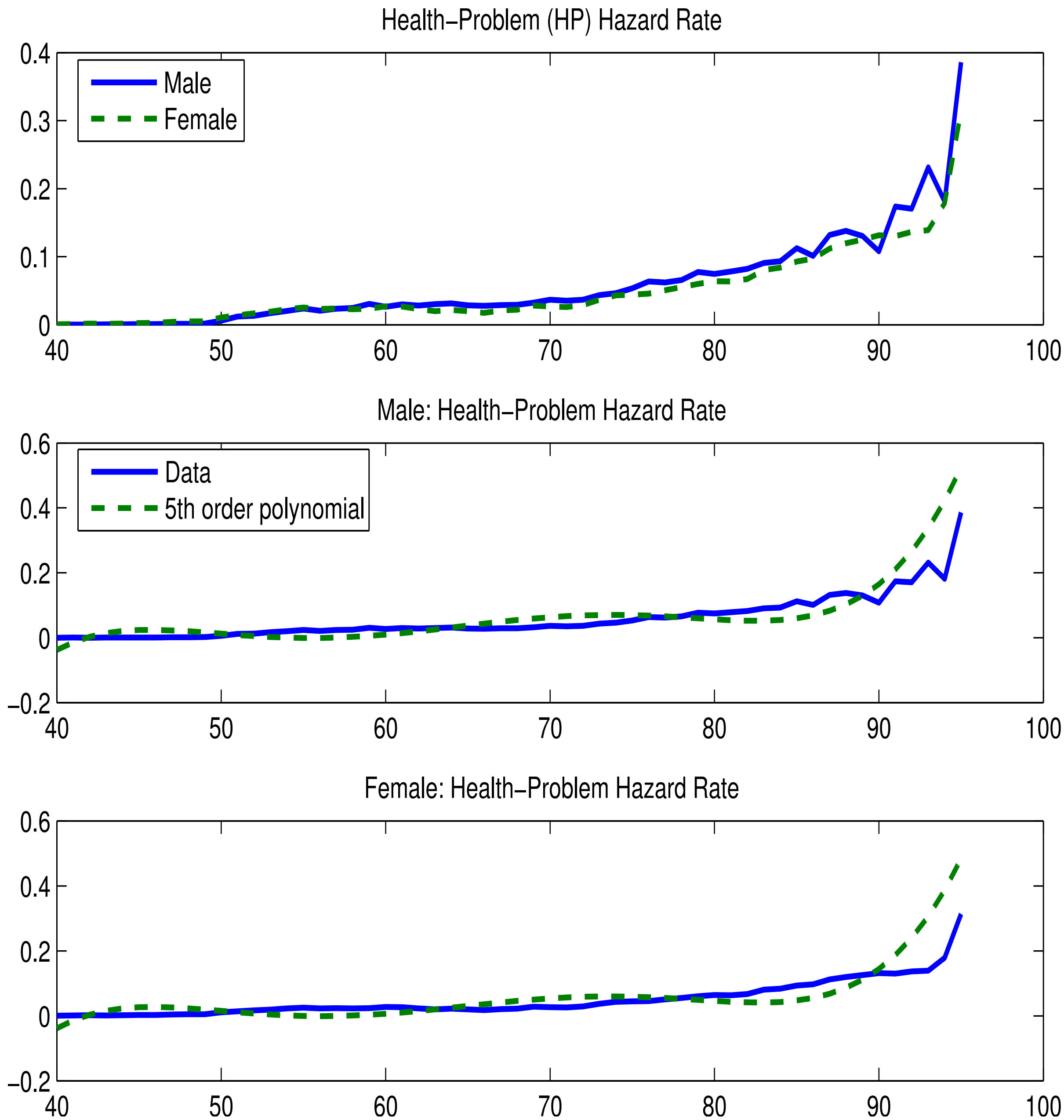


Figure 1: Work Limiting Health Problems Hazard Rate. Original Data from RAND-HRS,Wave 1-6. Fitted function is a 5th order polynomial, fitted with least squares.

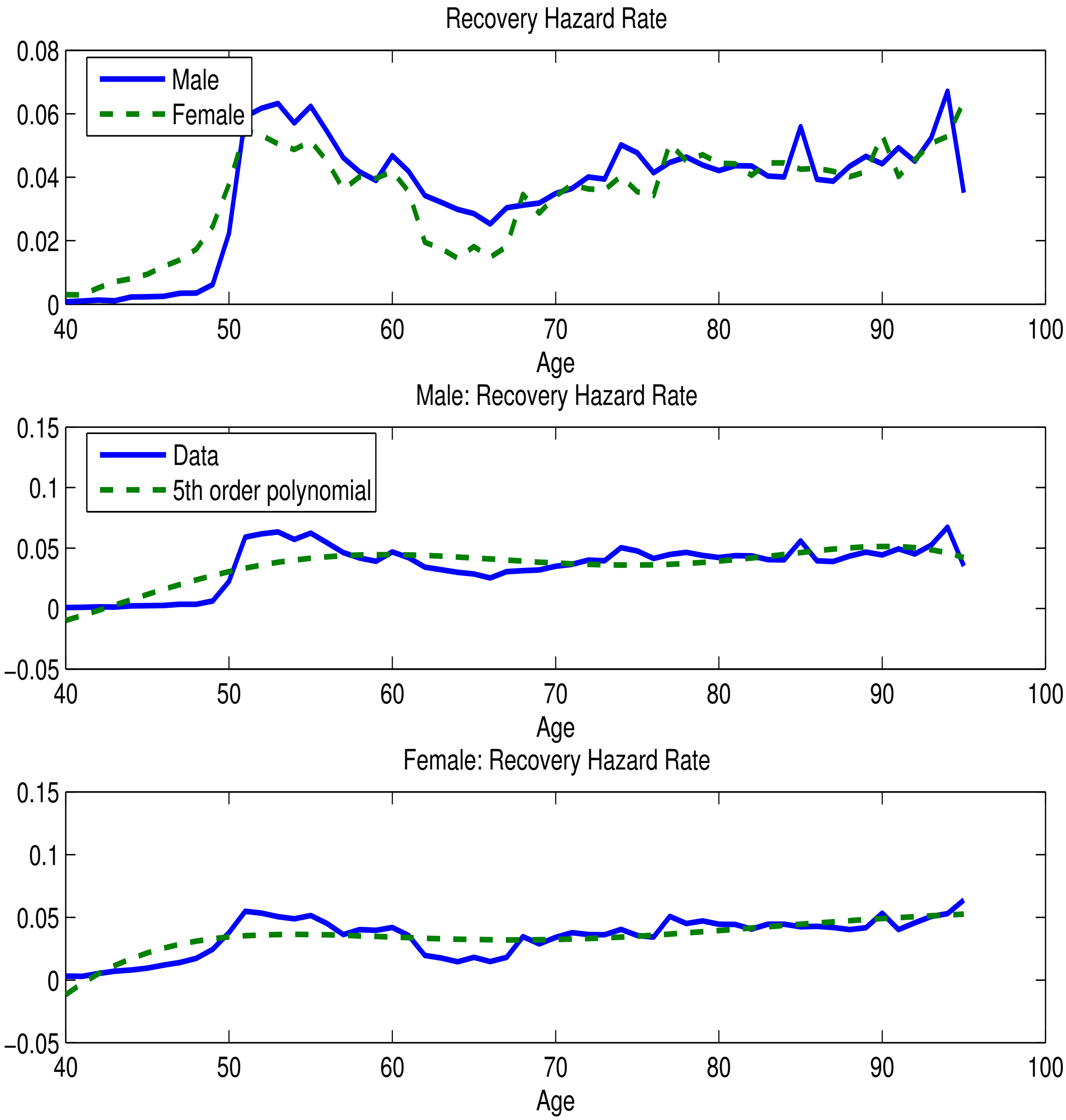
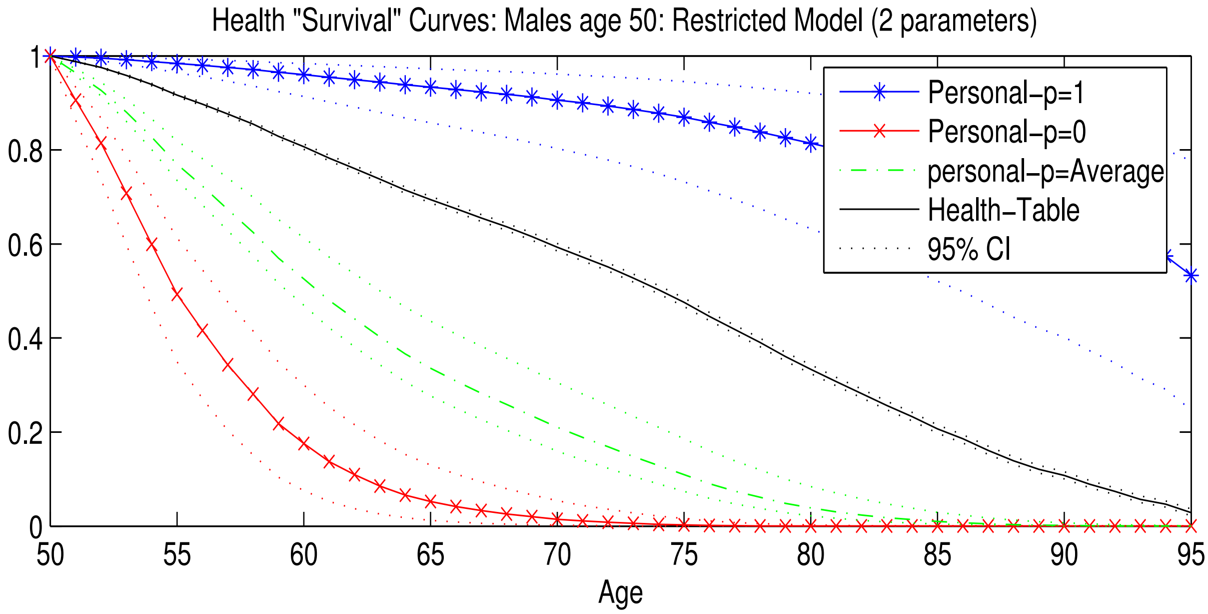


Figure 2: 'Recovery from Work Limiting Health Problems' Hazard Rate. Original Data from RAND-HRS,Wave 1-6. Fitted function is a 5th order polynomial, fitted with least squares.



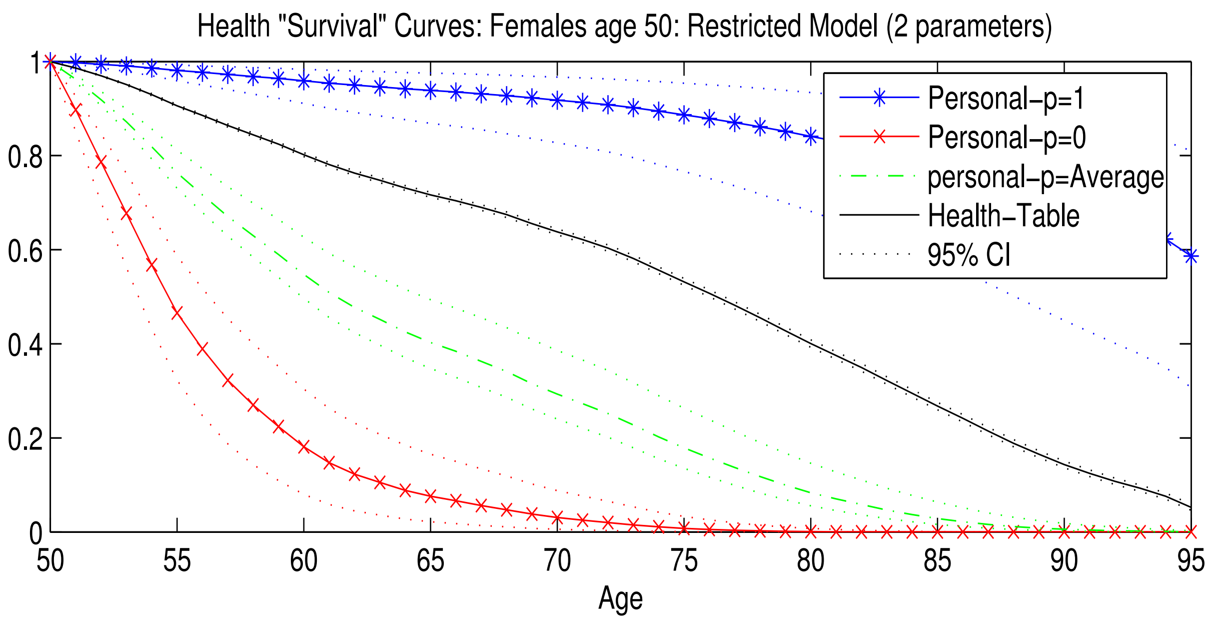


Figure 3: Health "Survival" Probabilites of a 50 Year Old.

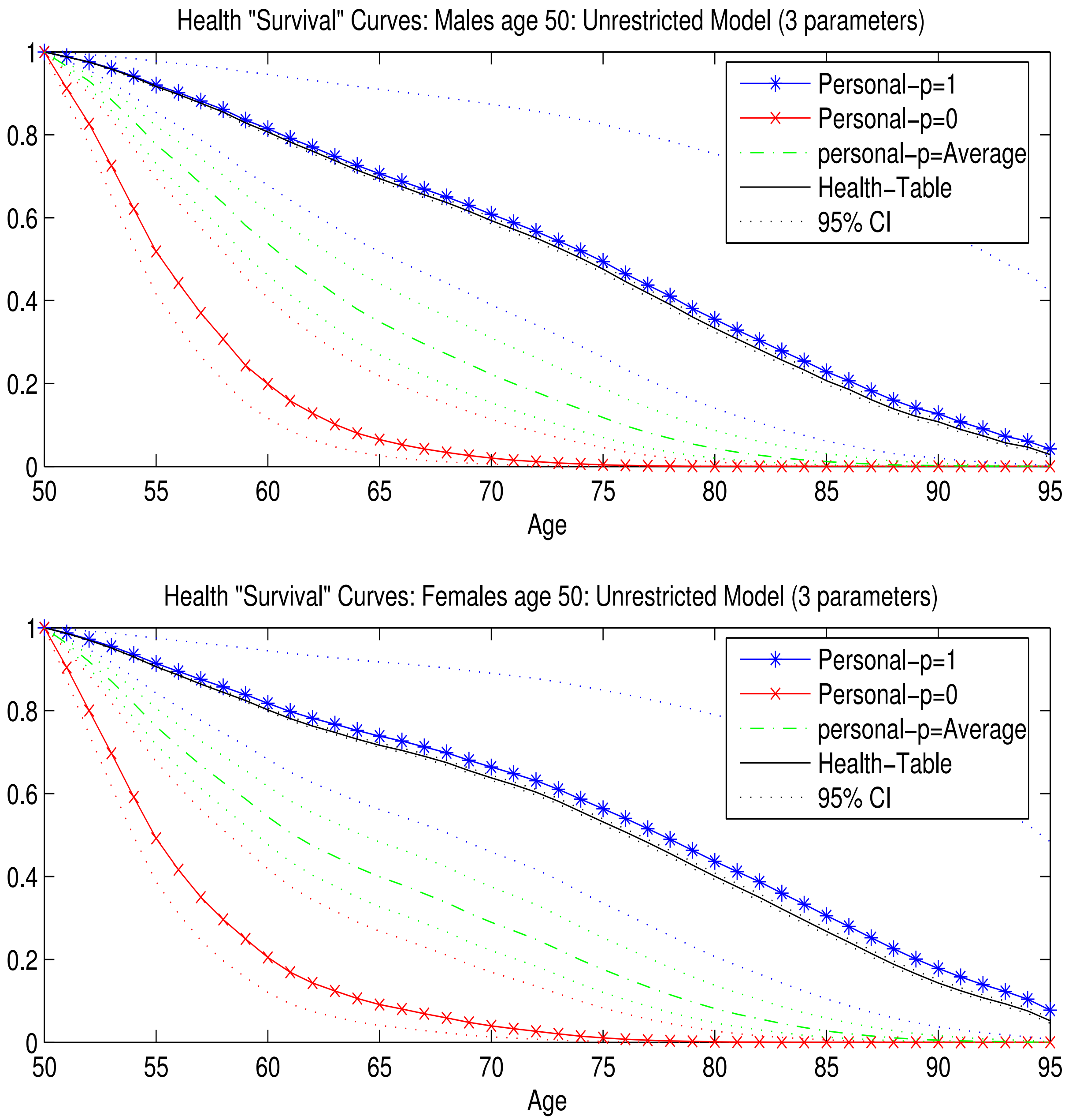
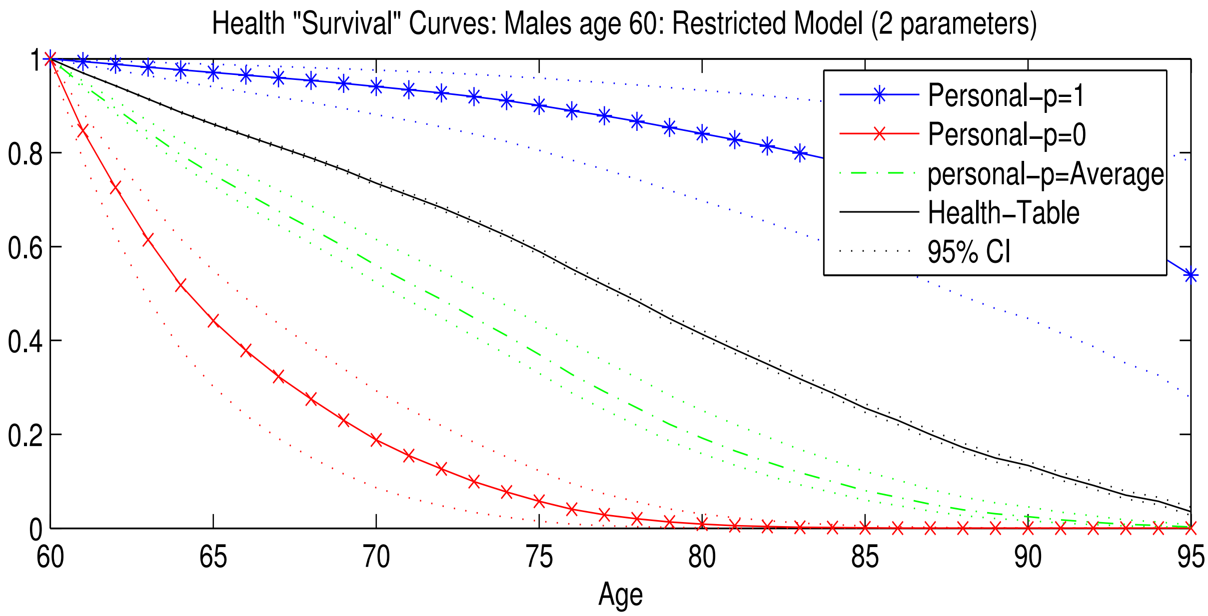


Figure 4: Health "Survival" Probabilites of a 50 Year Old for the Unrestricted Model.



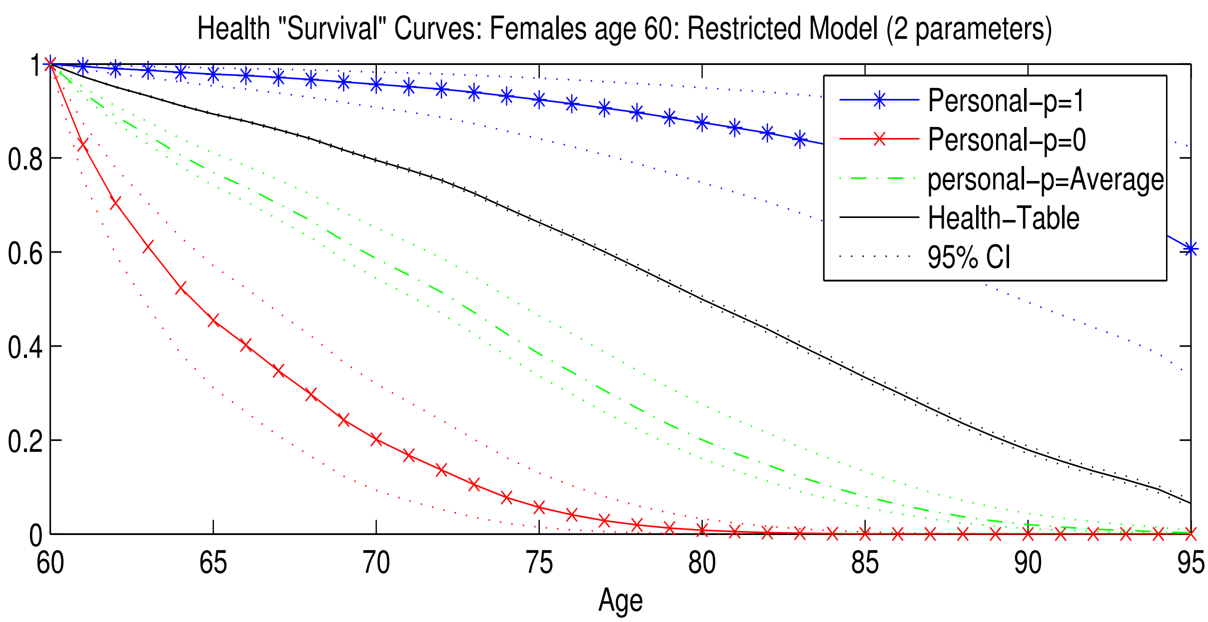
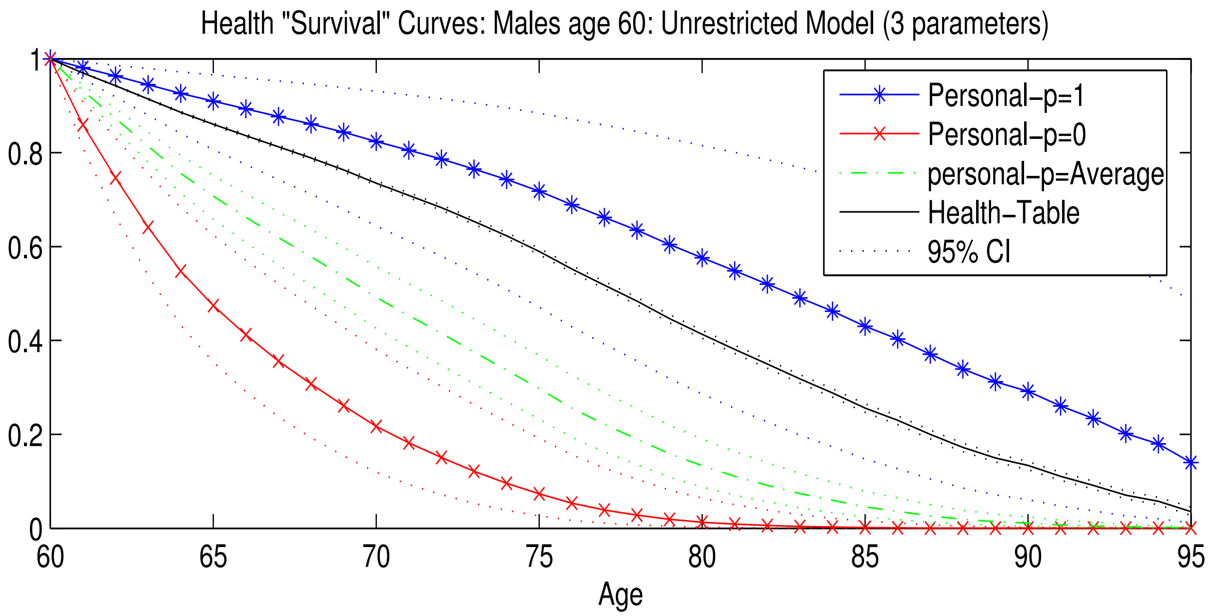


Figure 5: Health "Survival" Probabilites of a 60 Year Old.



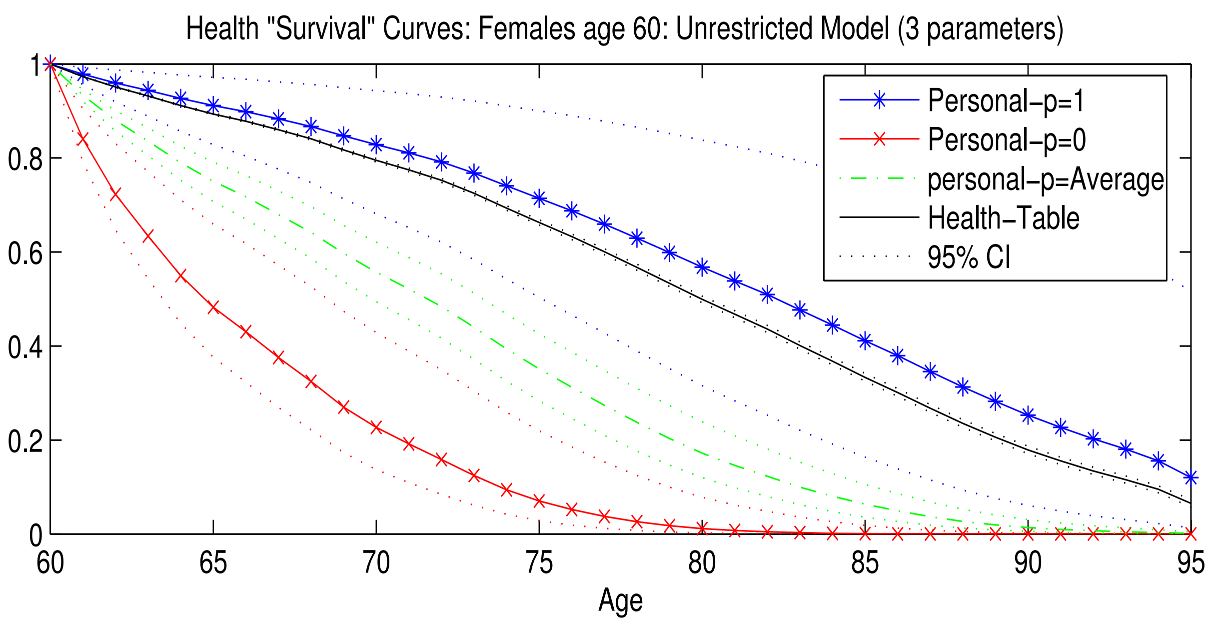


Figure 6: Health "Survival" Probabilites of a 60 Year Old for the Unrestricted Model.

$$\text{Prob(WorkLimHealthProblems}_j = 1) = \Phi(\beta ExpectedWorkLimHealthProblems_{92})$$

COEFFICIENT	(1) Wave2_mfx HealthProb94	(2) Wave3_mfx HealthProb96	(3) Wave4_mfx HealthProb98	(4) Wave5_mfx HealthProb00	(5) Wave6_mfx HealthProb02	(6) Wave7_mfx HealthProb04
ExpHealthProblem	0.00192***	0.00190***	0.00198***	0.00201***	0.00218***	0.000500***
	(0.00013)	(0.00015)	(0.00015)	(0.00016)	(0.00016)	(0.00012)
Observations	7238	7238	7238	7238	7238	7238

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

Table 10: Validity of Work Limiting Health Expectations from a Probit. We report marginal effects. Dependent variable is Work Limiting Health Problems in Wave 2,3,4,5,6, and 7. The regressor is expected work limiting health problems of wave1. Age: 40-60 in Wave 1.

$$\text{Prob(WorkLimHealthProblems}_j = 1) = \Phi(\beta ExpectedWorkLimHealthProblems_{92} + \gamma X_{92})$$

COEFFICIENT	(1) Wave2_mfx HealthProb94	(2) Wave3_mfx HealthProb96	(3) Wave4_mfx HealthProb98	(4) Wave5_mfx HealthProb00	(5) Wave6_mfx HealthProb02	(6) Wave7_mfx HealthProb04
ExpHealthProblem	0.000618***	0.000519***	0.000672***	0.000778***	0.000986***	0.000273**
	(0.00013)	(0.00015)	(0.00015)	(0.00016)	(0.00017)	(0.00013)
age	0.00138	0.00156	0.00363***	0.000911	0.00236**	-0.0000130
	(0.00094)	(0.0010)	(0.0011)	(0.0011)	(0.0012)	(0.00088)
female	-0.0381***	-0.0442***	-0.0255***	-0.0197**	-0.0167	-0.0291***
	(0.0080)	(0.0093)	(0.0094)	(0.010)	(0.010)	(0.0076)
black	-0.0103	-0.0243**	-0.0218**	-0.0156	-0.0123	-0.0214**
	(0.0090)	(0.010)	(0.011)	(0.012)	(0.012)	(0.0085)
partner	-0.0226**	-0.0211**	0.00423	-0.00939	-0.0178	-0.0229**
	(0.0096)	(0.011)	(0.010)	(0.012)	(0.012)	(0.0095)
Education > 12 years	0.000259	0.00436	-0.0118	-0.0101	-0.00653	-0.00978
	(0.0077)	(0.0088)	(0.0091)	(0.0097)	(0.010)	(0.0073)
Health status: Good	0.0944***	0.0882***	0.0936***	0.115***	0.109***	0.0379***
	(0.014)	(0.015)	(0.015)	(0.016)	(0.016)	(0.011)
Health status: Fair	0.214***	0.195***	0.189***	0.189***	0.156***	0.0281*
	(0.027)	(0.026)	(0.026)	(0.026)	(0.025)	(0.016)
DocDiag. stroke	0.0797*	0.0527	0.0226	0.0848*	0.0181	-0.0298
	(0.042)	(0.042)	(0.039)	(0.049)	(0.043)	(0.025)
DocDiag. arthritis/rheumatism	0.0297***	0.0563***	0.0594***	0.0534***	0.0739***	0.0121
	(0.0083)	(0.0097)	(0.010)	(0.011)	(0.011)	(0.0081)
Smokes now	0.0220**	0.0459***	0.0406***	0.0366***	0.0461***	0.0264***
	(0.0094)	(0.011)	(0.011)	(0.012)	(0.012)	(0.0095)
Individual earnings (in 1000)	-0.000884***	-0.000546**	-0.000589***	-0.000577**	-0.000684***	-0.000207
	(0.00022)	(0.00025)	(0.00022)	(0.00023)	(0.00023)	(0.00015)
Employed	-0.0883***	-0.0152	-0.0110	-0.00573	0.00577	-0.0375*
	(0.024)	(0.022)	(0.022)	(0.023)	(0.023)	(0.021)
Observations	7158	7158	7158	7158	7158	7158

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

Table 11: Validity of Work Limiting Health Expectations from a Probit. We report marginal effects. Dependent variable is Work Limiting Health Problems in Wave 2,3,4,5,6, and 7. All regressors are from Wave 1. Further explanatory variables include doctor diagnosed health problems, change in doctor diagnosed health problems, wealth and income variables, demographic variables and life style variables (these are not reported here). Age: 40-60 in Wave 1.

$$\text{WorkLimHealthProblems}_j = \beta \times ExpectedWorkLimHealthProblems_{92} + X_{92}\gamma + Z_{92}\theta$$

COEFFICIENT	(1) HealthProb94	(2) HealthProb96	(3) HealthProb98	(4) HealthProb00	(5) HealthProb02	(6) HealthProb04
ExpHealthProblem	0.00225	0.00766	0.00666	0.00411	0.00447	0.00462
	(0.0038)	(0.0049)	(0.0048)	(0.0046)	(0.0047)	(0.0037)
age	0.000575	-0.00163	0.000812	-0.000583	0.000553	-0.00194
	(0.0017)	(0.0022)	(0.0022)	(0.0021)	(0.0021)	(0.0017)
female	-0.0327***	-0.0305**	-0.0121	-0.0103	-0.00516	-0.0221**
	(0.0099)	(0.012)	(0.012)	(0.012)	(0.012)	(0.0097)
black	-0.0130	-0.0213	-0.0191	-0.0159	-0.0114	-0.0177
	(0.012)	(0.015)	(0.015)	(0.015)	(0.015)	(0.011)
partner	-0.0216**	-0.0153	0.0107	-0.00732	-0.0153	-0.0177
	(0.011)	(0.013)	(0.013)	(0.013)	(0.013)	(0.011)
Education > 12 years	-0.00200	0.00126	-0.0135	-0.0104	-0.00802	-0.0114
	(0.0079)	(0.0099)	(0.0098)	(0.0096)	(0.0099)	(0.0081)
Health status: Good	0.0451	-0.0168	0.00825	0.0565	0.0526	-0.0111
	(0.041)	(0.052)	(0.052)	(0.049)	(0.050)	(0.040)
Health status: Fair	0.157**	0.0418	0.0694	0.110	0.0773	-0.0513
	(0.070)	(0.089)	(0.088)	(0.083)	(0.086)	(0.068)
DocDiag. stroke	0.109**	0.0494	0.0135	0.0844	0.00837	-0.0519
	(0.052)	(0.056)	(0.055)	(0.055)	(0.054)	(0.036)
DocDiag. arthritis/rheumatism	0.0306**	0.0442***	0.0501***	0.0488***	0.0701***	0.00223
	(0.013)	(0.017)	(0.016)	(0.016)	(0.016)	(0.013)
Smokes now	0.0224*	0.0310*	0.0271	0.0294*	0.0377**	0.0167
	(0.013)	(0.017)	(0.017)	(0.016)	(0.016)	(0.013)
Individual earnings (in 1000)	-0.000397***	-0.000202	-0.000201	-0.000273*	-0.000300**	-0.0000610
	(0.00014)	(0.00017)	(0.00015)	(0.00016)	(0.00015)	(0.00013)
Employed	-0.110***	-0.0358	-0.0315	-0.0153	-0.00770	-0.0496**
	(0.025)	(0.028)	(0.027)	(0.027)	(0.027)	(0.023)
Constant	0.0200	-0.0610	-0.217***	-0.135*	-0.219***	0.0779
	(0.068)	(0.080)	(0.083)	(0.079)	(0.080)	(0.065)
Observations	7158	7158	7158	7158	7158	7158
R^2	0.14	-0.13	-0.05	0.06	0.05	-0.14
P-val Durbin-Wu-Hausman	0.717	0.118	0.206	0.504	0.496	0.273
P-val Hansen J-Stat	0.896	0.639	0.265	0.328	0.322	0.558
Cragg-Donald statistic	0.956	0.956	0.956	0.956	0.956	0.956

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 12: Validity of Work Limiting Health Expectations from Linear Probability IV-Regression. Dependent variable is Work Limiting Health Problems in Wave 2,3,4,5,6, and 7. The instrumented regressor is expected work limiting health problems of wave1. Instruments are 12 indicator variables constructed from age of parents when alive or at death. Sample Age: 40-60 in Wave 1.

$$\text{Prob(WorkLimHealthProblems}_j = 1) = \Phi(\beta \times ExpectedWorkLimHealthProblems_{92} + Z_{92}\theta)$$

COEFFICIENT	(1) Wave2_mfx	(2) Wave3_mfx	(3) Wave4_mfx	(4) Wave5_mfx	(5) Wave6_mfx	(6) Wave7_mfx
ExpHealthProblem	0.0312***	0.0306***	0.0413***	0.0382***	0.0390***	0.0172*
	(0.0096)	(0.0092)	(0.0099)	(0.0095)	(0.0094)	(0.0099)
Observations	7238	7238	7238	7238	7238	7238

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

Table 13: Validity of Work Limiting Health Expectations from IV-Probit Estimation. We report marginal effects. Dependent variable is Work Limiting Health Problems in Wave 2,3,4,5,6, and 7. The instrumented regressor is expected work limiting health problems of wave1. Instruments are 12 indicator variables age of parents when alive or at death. Sample Age: 40-60 in Wave 1.

$$\text{Prob(WorkLimHealthProblems}_j = 1) = \Phi(\beta \times ExpectedWorkLimHealthProblems_{92} + X_92\gamma + Z_{92}\theta)$$

COEFFICIENT	(1) Wave2 HealthProb94	(2) Wave3 HealthProb96	(3) Wave4 HealthProb98	(4) Wave5 HealthProb00	(5) Wave6 HealthProb02	(6) Wave7 HealthProb04
ExpHealthProblem	0.00389	0.00427	0.00305	0.00260	0.00560	0.00139
	(0.0034)	(0.0037)	(0.0037)	(0.0039)	(0.0042)	(0.0031)
age	-0.0000609	-0.000310	0.00221	0.00000213	0.000116	-0.000684
	(0.0016)	(0.0017)	(0.0018)	(0.0019)	(0.0020)	(0.0015)
female	-0.0304***	-0.0353***	-0.0172	-0.0124	-0.00357	-0.0267***
	(0.0098)	(0.011)	(0.011)	(0.011)	(0.012)	(0.0088)
black	-0.0109	-0.0257**	-0.0237*	-0.0178	-0.00997	-0.0218**
	(0.012)	(0.013)	(0.013)	(0.014)	(0.015)	(0.010)
partner	-0.0199*	-0.0186	0.00717	-0.00882	-0.0142	-0.0209**
	(0.011)	(0.012)	(0.012)	(0.012)	(0.013)	(0.010)
Education > 12 years	-0.00266	0.00263	-0.0121	-0.00984	-0.00847	-0.0101
	(0.0082)	(0.0090)	(0.0090)	(0.0093)	(0.010)	(0.0075)
Health status: Good	0.0280	0.0184	0.0457	0.0721*	0.0409	0.0224
	(0.037)	(0.040)	(0.041)	(0.042)	(0.046)	(0.033)
Health status: Fair	0.128**	0.102	0.133*	0.136*	0.0574	0.00588
	(0.063)	(0.070)	(0.069)	(0.073)	(0.077)	(0.057)
DocDiag. stroke	0.104**	0.0594	0.0241	0.0888*	0.00507	-0.0424
	(0.053)	(0.052)	(0.051)	(0.054)	(0.055)	(0.033)
DocDiag. arthritis/rheumatism	0.0266**	0.0526***	0.0589***	0.0525***	0.0673***	0.0101
	(0.013)	(0.014)	(0.014)	(0.015)	(0.015)	(0.012)
Smokes now	0.0187	0.0386***	0.0352**	0.0328**	0.0352**	0.0240**
	(0.013)	(0.014)	(0.014)	(0.015)	(0.016)	(0.012)
Individual earnings (in 1000)	-0.000362***	-0.000275*	-0.000278**	-0.000305**	-0.000276*	-0.000130
	(0.00013)	(0.00015)	(0.00014)	(0.00015)	(0.00015)	(0.00011)
Employed	-0.114***	-0.0279	-0.0231	-0.0117	-0.0103	-0.0421*
	(0.025)	(0.025)	(0.024)	(0.026)	(0.027)	(0.022)
Constant	0.0197	-0.0603	-0.216***	-0.135*	-0.219***	0.0784
	(0.070)	(0.072)	(0.077)	(0.077)	(0.082)	(0.061)
Observations	7158	7158	7158	7158	7158	7158
R^2	0.09	0.07	0.10	0.09	0.01	0.01

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 14: Validity of Work Limiting Health Expectations from IV-Probit-Twostep Estimation. We report marginal effects. Dependent variable is Work Limiting Health Problems in Wave 2,3,4,5,6, and 7. The instrumented regressor is expected work limiting health problems of wave1. Instruments are 12 indicator variables age of parents when alive or at death. Sample Age: 40-60 in Wave 1. All regressors are from Wave 1. Further explanatory variables include doctor diagnosed health problems, change in doctor diagnosed health problems, wealth and income variables, demographic variables and life style variables (not all are reported here). Sample Age: 40-60 in Wave 1.

$$\text{ExpHealthProblems}_{i,t+1} = \alpha_i + \beta \times ExpHealthProblems_{it} + \Omega_{it}\gamma + \epsilon_{it}$$

	strongRE	weakRE	IV-strongRE	IV-weakRE
	(1)	(2)	(3)	(4)
Adjust-Health-Expectation	.286***	.286***	.923***	.923***
cons	30.961***	30.961***	4.866***	4.866***
e(N)	19866	19866	15324	15324
e(r2)				

Table 15: Tests for Weak and Strong Rationality using a Linear Probability Model. We use entire Panel 1992-2002.

Mincer-Type Regression: Log(wage$_{it}$) $= \beta \times WorkLimHealthProblems_{it} + x'_{it}\gamma + \epsilon_{it}$	poolOLS	poolOLSr	OLSpanel	Betweeniid	FixEffiid	RandomEiid
	(1)	(2)	(3)	(4)	(5)	(6)
HealthProblems	-.193	-.193	-.193	-.287	-.020	-.089
	(.020)***	(.022)***	(.026)***	(.036)***	(.021)	(.018)***
Employed	.241	.241	.241	.211	.279	.266
	(.024)***	(.032)***	(.038)***	(.045)***	(.025)***	(.022)***
Job requires physical effort	-.107	-.107	-.107	-.135	.00008	-.059
	(.014)***	(.013)***	(.017)***	(.024)***	(.019)	(.015)***
Hours worked	.062	.062	.062	.066	.054	.058
	(.001)***	(.003)***	(.003)***	(.003)***	(.002)***	(.001)***
Squared hours worked	-.0004	-.0004	-.0004	-.0005	-.0004	-.0004
	(.00002)***	(.00004)***	(.00004)***	(.00003)***	(.00002)***	(1.00e-05)***
Years worked	.029	.029	.029	.027	.059	.024
	(.002)***	(.003)***	(.003)***	(.004)***	(.013)***	(.003)***
Years worked2	-.0004	-.0004	-.0004	-.0004	-.0002	-.0003
	(.00004)***	(.00005)***	(.00006)***	(.00006)***	(.00009)**	(.00005)***
Age	.010	.010	.010	-.003	.010	.044
	(.010)	(.011)	(.015)	(.015)	(.022)	(.012)***
Age2	-.0001	-.0001	-.0001	-.00005	-.0002	-.0003
	(.00008)	(.0001)	(.0001)	(.0001)	(.0002)	(.0001)***
Male	-.290	-.290	-.290	-.320		-.264
	(.015)***	(.015)***	(.022)***	(.025)***		(.023)***
Education > 12 years	.080	.080	.080	.033		.203
	(.111)	(.121)	(.160)	(.178)		(.133)
Age*Education>12	.003	.003	.003	.003	.0009	.002
	(.002)	(.002)	(.003)	(.003)	(.003)	(.002)
Male*Education>12	.043	.043	.043	.037		.047
	(.017)**	(.016)***	(.024)*	(.028)		(.027)*
Female*Education>12	.059	.059	.059	.055		.074
	(.017)***	(.016)***	(.024)**	(.027)**		(.027)***
Constant	3.714	3.714	3.714	3.622	3.238	2.191
	(.286)***	(.325)***	(.421)***	(83326.350)	(.620)***	(.648)***
Number of observations	14019	14019	14019	14019	14019	14019
R^2	.495	.495	.495	.534	.264	
F-statistic	350.732	.	.	135.744	94.264	

Table 16: Linear Panel Wave (1-6): Dependent Variable is log(Wage)

$$\text{Prob(WorkLimHealthProblems}_6 = 1) = \frac{exp(\beta \times ExpectedWorkLimHealthProblems + X\gamma)}{1+exp(\beta \times ExpectedWorkLimHealthProblems + X\gamma)}$$

	Logit1	Logit2	Logit3	Logit4	Logit5	Logit6
	(1)	(2)	(3)	(4)	(5)	(6)
ExpHealthProblem	.004 (.002)**	.004 (.002)**	.004 (.002)**	.003 (.002)*		
ExpLive to 75	.004 (.003)		-.0007 (.002)		-.002 (.001)	
ExpLive to 85	-.006 (.002)***			-.004 (.002)**		-.004 (.001)***
Health status: Very Good	.275 (.157)*	.232 (.149)	.303 (.156)*	.207 (.152)	.325 (.142)**	.275 (.138)**
Health status: Good	.689 (.160)***	.738 (.150)***	.776 (.158)***	.646 (.155)***	.808 (.142)***	.707 (.139)***
Health status: Fair	.827 (.206)***	.826 (.188)***	.854 (.199)***	.750 (.198)***	.912 (.172)***	.772 (.171)***
Health status: Poor	.766 (.362)**	.811 (.312)***	.938 (.331)***	.739 (.354)**	.925 (.234)***	.705 (.241)***
Diff. walking accross room	-.773 (.525)	-.568 (.496)	-.628 (.501)	-.842 (.520)	-.180 (.302)	-.134 (.321)
Diff. walking 1 block	.421 (.174)**	.378 (.159)**	.404 (.163)**	.433 (.172)**	.616 (.133)***	.640 (.138)***
Diff. pushing large objects	.292 (.158)*	.246 (.146)*	.258 (.151)*	.244 (.155)	.355 (.123)***	.384 (.125)***
Diff. sitting 2 hours	.241 (.132)*	.250 (.122)**	.220 (.127)*	.291 (.128)**	.206 (.106)*	.268 (.108)**
Diff. using the phone	.605 (.468)	.352 (.383)	.441 (.430)	.401 (.428)	.325 (.315)	.402 (.339)
Diff. using money	.104 (.236)	-.004 (.223)	-.007 (.230)	.073 (.233)	-.229 (.198)	-.118 (.205)
Diff. climbing stairs	.396 (.119)***	.432 (.111)***	.432 (.115)***	.418 (.116)***	.359 (.101)***	.329 (.101)***
Diff. lifting 10 pounds	.337 (.158)**	.320 (.147)**	.366 (.151)**	.322 (.156)**	.423 (.123)***	.359 (.127)***
Felt depressed	-.017 (.146)	-.014 (.133)	-.007 (.138)	-.022 (.143)	-.013 (.116)	-.058 (.121)
Back problems	.253 (.111)**	.277 (.104)***	.275 (.107)**	.272 (.108)**	.250 (.093)***	.262 (.093)***
Constant	-5.586 (.868)***	-5.475 (.779)***	-5.248 (.833)***	-5.546 (.813)***	-4.121 (.715)***	-4.607 (.693)***
Number of observations	3380	3707	3538	3496	4399	4342

Table 17: Information Content of worklm, dependent variable is Work Limiting Health Problems in Wave 6. Further explanatory variables include doctor diagnosed health problems, change in doctor diagnosed health problems, wealth and income variables, demographic variables and life style variables (these are not reported here).