Estimating Markov Transition Probabilities between Health States in the HRS Dataset

Juergen Jung*
Indiana University - Bloomington

12th December 2006

Abstract

We estimate Markov transition probabilities for individual health status over time as function of observable characteristics. We implement 3 methods to construct these Markov probabilities. The first method is a counting method, the second predicts transition probabilities using ordered logit and ordered probit regression models and finally, we derive hazard rates from a non-parametric Kaplan-Meier estimator and a semi-parametric proportional hazard (Cox) model. We also test whether the Markov assumption holds. The estimated Markov switching matrix can be used in life-cycle models with health uncertainty.

1 Introduction

In this paper we estimate Markov transition probabilities between health states using the RAND-HRS dataset. These transition probabilities are important in life-cycle models with health uncertainty. Health uncertainty models have become more complex and do incorporate agents that are heterogenous in many dimensions (see Palumbo (1999), French (2003), French and Jones (2004), and De Nardi, French and Jones (2006)). Health and survival probabilities are often modeled as a function of age exclusively.

In addition, Markov switching models make the assumption that tomorrow's health state is a function of today's health state only. Since health is a complex variable that depends on many more factors than just age it is important to condition health transition probabilities (as well as mortality probabilities) on additional factors such as income, gender, education, life-style choices etc. Even if we maintain the Markov assumption, model predictions will be more accurate if different types of agents are modelled with a type-specific Markov switching matrix between health states.

The HRS is a suitable dataset for the construction of health transition probabilities since it reports five possible outcomes for its health status variable: 1. Excellent, 2. Very Good, 3. Good, 4. Fair, or 5. Poor. In addition, agents face mortality risk. We include the sixth state of death and use three separate methods to calculate Markov type health transition probabilities concentrating on the population of $40 - 90$ year olds.

The first method is nonparametric and involves a simple counting procedure, the second method uses predictions of ordered logit and ordered probit specifications and the third method combines elements of survival analysis. As a result we can condition these matrices on various individual characteristics like gender, age , income life-style, employment status etc. We plan to use these estimates in a follow up paper on health savings accounts with heterogenous agents and health uncertainty.

We next briefly describe the three methods we are using and the related literature. The first method to estimate transition probabilities between health states is a simple count method as

*I would like to thank Pravin Trivedi and Gerhard Glomm for many helpful discussions. Corresponding author: Juergen Jung, Indiana University - Bloomington, Tel.: 1-812-345-9182; *E-mail:* juejung@indiana.edu

used in Diehr et al. (1998) and Diehr and Patrick (2001). We count the number of transitions from initial states to states two periods ahead and group these counts by gender and age brackets.

The second method estimates an ordered logit specification with the health indicator variable as the dependent variable (augmented with death as the additional (absorptive) state) and a list of exogenous regressors as explanatory variables. We also run an ordered probit specification and a multinomial logit model on an extended list of regressors and compare the results. This method is widely applied in the life-cycle literature. Also the simpler linear probability model is used regularly. French (2003) and French and Jones (2004) use a linear probability model to estimate transition probabilities from two possible health states, good and bad using the PSID data and the HRS data respectively. In their model they estimate

$$\begin{aligned} & P\left(M_{it} = good|\alpha_i, M_{it-1}, age_{it}\right) \\ = \; & \alpha_i + \sum_{k=0}^{K} \beta_k age_{it}^k P\left(M_{it-1} = good|M_{it-1}^*\right) + \sum_{k=1}^{K} \gamma_k age_{it}^k P\left(M_{it-1} = bad|M_{it-1}^*\right) + \epsilon_{it}, \quad (1) \end{aligned}$$

where α_i represents individual heterogeneity in capacity for good health. De Nardi, French and Jones (2006) uses a somewhat different method and the AHEAD dataset. In order to find the one year transition rates of different health states, they first estimate the two year Markov transition matrix $P_{t+2|t}$ using a linear probability model.[1] Then they assume that the one-year Markov transition matrix, $P_{t+1|t}$, satisfies $P_{t+2|t} = P_{t+1|t}^2$, so that $P_{t+1|t}$ can then be found as the solution to a quadratic form. They find the somewhat surprising result that the probability of being in bad health this year, conditional on being in bad health last year, falls with age. However, this measures the probability of still being in bad health and surviving, conditional on being in bad health last period. With this in mind they find that the probability of either being dead or in bad health this year, conditional on being in bad health last year, remains constant at about 90% at each age. Palumbo (1999) estimates Markov transition probabilities for household health status over time as functions of observable household characteristics using a multinomial logit model applied to a sample of elderly families from the PSID.

Finally, we use methods from survival analysis and construct hazard rates from the various health states into consecutive health states 2 years later by means of non-parametric and semi-parametric methods. This method adjusts for censoring in the data since we have many individuals leaving the survey not due to death. Anderson, Hansen and Keiding (1991) present non- and semi-parametric methods to estimate transition probabilities from censored data and illustrate their method in a study on liver disease. More evolved methods of this kind can be found in Dabrowska, Sun and Horowitz (1994) and Cook, Lawless and Lee (2003). In those studies additional dependencies are introduced, e.g. covariates depending on backward recurrence time and adaptive censoring in Markov switching models with renewable states.

The paper is structured as follows. The next section describes the data. Section 3 introduces the simple counting method. Section 4 contains results from the ordered logit, ordered probit and multinomial logit regression models. Section 5 describes the non-parametric and semi-parametric survival approach. We conclude our analysis in section 6. Tables and figures are summarized in the appendix.

2 Data

We use seven waves of the RAND-HRS survey with surveys conducted every two years from 1992 to 2004. The RAND-HRS is developed from the health and retirement study (HRS) by the RAND Center of Aging. It is a composite data set that combines four cohort studies to get a national representative of the older population in the U.S. The cohorts are the AHEAD cohorts born before 1924, the CODA cohorts born between $1924-1930$, the HRS cohorts born between $1931-1941$ and the War Baby cohorts born between $1942-1947$. The largest of these

[1] The AHEAD dataset surveys individuals every two years.

surveys is the Health and Retirement Survey (HRS) collected by the Institute for Social Research at the University of Michigan. It is a longitudinal survey conducted every two years starting from 1992. Wave 7 for year 2004 has become available in October 2006 an is included in this project. The RAND-HRS data covers a broad range of topics, including health, income, assets, employment, retirement, insurance, and family structure.

Survey respondents are non-institutionalized individuals. The majority of them were between 51 and 61 years old when the survey was first conducted in 1992. The baseline survey included 12,652 persons, or 7,600 households, with over sampling of Mexican Americans, African Americans and residents of Florida. Juster and Suzman (1995) present a general overview of the HRS and Wallace and Herzog (1995) review the health measures in particular. Figure 1 contains histograms of the age distribution of all waves including a histogram of the age distribution over all waves. We see that the sample covers mostly individuals from age $45-75$. In the following we will concentrate on the age groups between 40 and 90 years.

The HRS includes five possible outcomes for self-reported health status: 1 Excellent, 2 Very Good, 3 Good, 4 Fair, or 5 Poor. In addition, agents face mortality risk. Table 1 reports the number of observations per wave including the number of reported deaths. Sample entries and exits other than deaths are not shown. Whenever an agent leaves the survey due to death it is reported. It is also reported whether a non-respondent has died or not. Using this information we construct a sixth health status state, denoted *Dead*. All other individuals leaving the survey are assumed to be alive with unknown health state. We treat them as censored observations in the survival analysis part but ignore censoring in the counting method as well as the logit/probit specifications. This and the fact that only non-institutionalized individuals take part in the survey can potentially skew our resuls towards more positive estimates of health transition probabilities.

To correct for some missing values in the health measure we interpolate health measures whenever there was a valid measurement before and after the missing observation (1.4% of all the data). That is, we replaced the missing value with the mean of the measures before and after. Mundahl (1998) finds that this method is less biased than other methods. Diehr and Patrick (2001) find that interpolated values where underdispersed and they therefore add a small random error. We follow their method and add a small amount of random error after interpolation and then round to the nearest health category.

We also use the following covariates in the regression frameworks that follow: age, age squared ($age2$), body mass index (bmi), a dummy for smoker ($smoken$), a dummy for female ($female$), a dummy for whether the individual lives with a partner ($partner$), a dummy for education of more than 12 years ($deducM12$), household total income in \$1000 ($hatota1000$) and total household income squared ($sqhatota1000$). Table 2 reports summary statistics on these variables.[2]

3 Counting

The first method to calculate transition probabilities between health states is a counting method. Since we have a panel over 12 years with observations every two years, we can simply count health transitions. Table 3 presents a summary of self reported health per age group if last periods health state has been reported. This is in effect the number of transitions in the data. An individual can contribute up to 6 transitions. Note that an individual who reports *Excellent* health in all 7 waves contributes 6 transitions of the type *Excellent* → *Excellent*. Table 4 groups these transitions by age and gender. The data contains more transitions for women than for men in almost all age cohorts.

We next count the realization of a particular transition as follows. Define the random variable of a health state realization at time t as $Y(t)$ where the realizations of this variable is $y(t) \in \{1,2,3,4,5,6\}$. Next we define the realization of a particular transition from state h to state j as

$$N_{hj} = \sum_{i=1}^{N} \sum_{t=1994}^{T=2004} I\left(y_i(t) = j, y_i(t-2) = h\right),$$

[2] It will turn out that our regression model will be misspecified due to omitted variables. Future version of this paper will include more covariates to describe health transitions.

where i counts for all individuals and t counts for the time periods from 1994, 1996, ..., 2004 in two year increments and $y_i(t)$ is the realization of the health state of individual i in year t. Thus N_{hj} registers the direct transition from $h \rightarrow j$ over all ages. Then the (average) transition probability from state h to j is denoted P_{hj} and calculated as

$$P_{hj}(t) = P(Y_t = h | Y_{t-1} = j) = \frac{N_{hj}}{\sum_{n=1}^{N} \sum_{t=1992}^{T} I(y_{it-1} = h)}.$$

We report these probabilities in table 5 for females and table 6 for males. The top panels in both tables are the transition probabilities, the bottom panels are the actual frequencies (or counts) that the probabilities are based on.

4 Predicted Markov Switching Probabilities from an Ordered Logit Model

The second method is based on Palumbo (1999). First we assume that health states can only obtain five possible outcomes: excellent $(h_{it} = 1)$, very good $(h_{it} = 2)$, good $(h_{it} = 3)$, fair $(h_{it} = 4)$, or poor $(h_{it} = 5)$. We denote $p_i(k|j)$ the probability that household i draws health status k in period t, conditional on having drawn health status j in period $t-1$. In addition, we allow transition probabilities to vary with several household characteristics: age, age^2, body mass index (bmi), smoker ($smoken$), partnership status ($partner$), gender ($female$), income ($hatota1000$), income^2 ($sqhatota1000$) and education ($deducM12$). We summarize these characteristics for household i in period t in vector x_{it},

$$x_{it} = \left\{x_{age,t},\ x_{age2,t},\ x_{female},\ x_{partner,t},\ x_{deducM12},\ x_{rbmi,t},\ x_{rsmoken,t},\ x_{hatota1000,t},\ x_{sqhatota1000,t}\right\}_i.$$

Next we define the index variable $v_i(k|j)$ as

$$v_i(k|j) = \begin{cases} \exp\left(x_{it}'\beta_{kj}\right), & \text{for } k = 1, 2, 3, \text{ or } 4, \\ \exp\left(x_{it}'0_{5j}\right), & \text{for } k = 5 \text{ (we normalize } \beta_5 = 0). \end{cases} \tag{2}$$

Then we define the Markov transition probabilities for the health states of household i in year t as

$$p_i(k|j) = \frac{v_i(k|j)}{\sum_{j=1}^{5} v_i(k|j)},$$

which is the multinomial logit specification. We will estimate a version of this later. Since there is a natural ordering in the health states we can put more structure on the model. Using the latent variable

$$y_{it}^* = x_{it}'\beta + u_{it}.$$

The latent variable crosses a series of thresholds, which determines the ordering. When for example $y^* > \alpha_1$ then the health status improves from *Poor* to $Fair$, if $y^* > \alpha_2$ health improves to *Good* etc.[3] It can then be shown that for an $m - alternative$ ordered model we get

$$\begin{aligned} \Pr[y_i = j] &= F(\alpha_j - x_i'\beta) - F(\alpha_{j-1} - x_i'\beta), \\ &= \frac{\exp(\alpha_j - x_i'\beta)}{1 + \exp(\alpha_j - x_i'\beta)} - \frac{\exp(\alpha_{j-1} - x_i'\beta)}{1 + \exp(\alpha_{j-1} - x_i'\beta)}, \end{aligned}$$

[3] In our data the ordering is actually reverse, so 1.*Excellent*, 2.$Very$ *Good*, 3.*Good*, ... insteady of 1.*Poor*, 2.$Fair$, 3.*Good*, ... This ordering does not matter as the marginal effects and predictions from the ordered logit are done correctly for each target group. The direction of ordering is irrelevant.

so that the log likelihood function is

$$\begin{aligned}\ln L(\beta,\alpha) &= 1[y_i=1]\log\left[\frac{\exp(\alpha_1-x_i'\beta)}{1+\exp(\alpha_1-x_i'\beta)}\right]+\\ & 1[y_i=2]\log\left[\frac{\exp(\alpha_2-x_i'\beta)}{1+\exp(\alpha_2-x_i'\beta)}-\frac{\exp(\alpha_1-x_i'\beta)}{1+\exp(\alpha_1-x_i'\beta)}\right]+\\ & \vdots \\ & 1[y_i=5]\log\left[1-\frac{\exp(\alpha_5-x_i'\beta)}{1+\exp(\alpha_5-x_i'\beta)}\right].\end{aligned}$$

This function is well behaved and can be estimated for β and α. Estimates of $\hat{\beta}$ are estimated relative to the health transition into poor health because $\hat{\beta}_{5j}$ is normalized to 0 for $j=1,2,3,4$, and 5.[4]

We then estimate the coefficients in (2) for five separate ordered logit models – one for each of the five health states being conditioned upon $j=1,2,3,4$ and 5. In this sense we run five separate ordered logit models on five different data sets. The first data set includes the current health status and household characteristics for all families in excellent health in the previous year. The second, third, fourth and fifth data sets include families in very good, good, fair and poor health in previous years respectively.

Next consider the estimation of vector $\hat{\beta}_{k1}$ for $k=1,2,3,4$, and 5, that is, the effects of household characteristics on health transition fitted values, conditional on a family having been in excellent health last year. The data we use to estimate these coefficient vectors consist of:

1. 1994 health status and 1994 household characteristics for families in excellent health in 1992,
2. 1996 health status and household characteristics for families in excellent health in 1994,
3. 1998 health status and household characteristics for families in excellent health in 1996,
4. 2000 health status and household characteristics for families in excellent health in 1998, and finally
5. 2002 health status and household characteristics for families in excellent health in 2000.

We use this model to predict the probabilities to be in the 5 health states respectively. Since these predictions are conditioned on last periods health state, we get predictions for each individual conditional on the previous health state.[5] We then report mean values of the predictions conditioned on the previous health states to fill the 5×5 Markov transition matrix. The mean is over all individuals and all age groups from $40-90$ years old. We report separate Markov switching matrices for women and men in tables 7 and 8 respectively.

Note that these probabilities are conditioned on survival since only surviving individuals add observations to the regression model. In order to make the predictions comparable to the counting method in section 3 we augment the health state with an additional sixth state: *death*. This is an absorptive state. The health indicator variable is now defined as: excellent $(h_{it}=1)$, very good $(h_{it}=2)$, good $(h_{it}=3)$, fair $(h_{it}=4)$, poor $(h_{it}=5)$, or dead $(h_{it}=6)$.

Since individuals do not report time varying data like household income ($hatota1000$), squared household income ($sqhatota1000$), body mass index (bmi), smoking statues ($smoken$), and partnership status ($partner$) in the event of death we have to augment the data for those individuals who died. The assumption that we make is that household income and as well as all other time varying variables of a deceased individual are the same as at the time of the last interview two

[4] Palumbo (1999) uses a multinomial logit model. Since we only report means of the predicted values, the results do not change between an ordered logit or multinomial logit model.

[5] Alternatively we could use a lag dummy for last period's health status in the above regression and set this variable to the according health states in order to condition the predictions.

years prior. The maximum informational lag that we introduce with this method is therefore 2 years.[6]

We report the mean of individual predictions over the age range from $40-90$ year olds with the state of death included in table 9 for women and in table 10 for men. The probabilities in these two tables are not conditional on survival since survival is an element of the transition matrices.

If we compare these unconditional health transitions with the health transitions conditional on survival (reported previously in tables 7 and 8) we see that the transition probabilities from poor to poor are much larger in the cases conditional upon survival, whereas the unconditional survival probabilities account for the fact that many individuals in poor health will transition to state 6, death. It therefore makes sense that the $6-state$ Markov tables report lower probabilities to stay in poor health.

We report marginal effects of the ordered logit model in table 7.1 and 7.1. Marginal effects are evaluated for a 50 year old individual at the median of household income and the mean of body mass index. The categorical variables are evaluated at $female = 0$, $partner = 1$, $deducM12 = 1$ and $rsmoken = 0$).

All regressors are highly significant except the variable indicating partnership (married or unmarried). Also, if the initial health state is *Poor*, regressors capturing partnership status, education status, household total income, and smoking become less significant. This also has to do with the smaller sample size for that particular health group. We report cluster robust estimates with a cluster being defined by the individual. This accounts for the fact that an individual can contribute up to 6 transitions, depending on how long they are in the survey.

The signs of the marginal effect confirm our intuition that women have better health prospects than men. They are more likely to stay in good health or to transition towards good health if the initial health state is good.

Smoking has a negative effect on transition probabilities to either stay or move towards good health. Body mass index is also negative related to the probability of remaining in *Excellent* health. However, if the initial health state is either *Fair* or *Poor* then the body mass index has a positive effect on the probability of transitioning towards better health states. The later observation makes sense, since among the group of individuals with *Poor* health, those with higher weights are probably the ones less weakened by sickness.

We are also interested in how the survival probabilities change with respect to age. Figures 2, 3, 4, 5 and 6 report transition probabilities as a function of age and gender. These probabilities are not means of predicted probabilities for the population like in the tables before. The probabilities in the figures are predictions evaluated at the same values as the marginal effects, that is at the median of household income and the mean of body mass index. The categorical variables are evaluated at $partner = 1$, $deducM12 = 1$ and $rsmoken = 0$).

We also report the 95% confidence intervals and see that the predictions for males and females overlap in most cases with exceptions in the age ranges close to age 40 and close to age 90. This is due to the smaller sample size in those age ranges. Compare again figure 1 for the age distribution of our sample.

From the figures we can see that transition probabilities to the good health states like *Excellent* and *Very Good* are declining functions of age, whereas transition probabilities to the bad health states like *Poor*, and *Dead* are increasing functions of age. This makes intuitive sense as one would expect that health deteriorates as a function of age and transitions to the better health states should become less likely, whereas transitions to worse health states are more probable.

[6] We could alternatively augment the data using

$$\widehat{income}_{year \text{ of death}} = f\left(income_{year\ of\ death-2}, income_{year\ of\ death-4}, \ldots\right) + \varepsilon_t,$$

where *year* of death is the survey year when the individual is reported dead. The year of actual death could be up to 2 years earlier. Function f could be a simple extrapolation from past trends in household incomes with some random component added. We have not experimented with any specific functional forms yet but plan do to so in the future.

Transition probabilities to the intermediate health states *Good* and $Fair$ are increasing functions of age when the previous health state was better than *Good* or $Fair$. If the previous health state was worse than *Good* or $Fair$, then the respective transition probabilities are declining functions of age. Both these observations are again consistent with the idea that health is likely to deteriorate as an individual gets older.

To check for robustness of our model specification we also report estimation results from an ordered probit model in tables 11 and 12 for the $6-state$ Markov model. The results are almost identical to the ordered logit model. Figures 7, 8, and 9 report the same transition probabilities with respect to age for an ordered logit and an ordered probit model. Since the estimation results have already been reported to be very similar between the two models, we again confirm that the confidence intervals of the graphs overlap. For brevity we only report the transitions from *Excellent*, *Good*, and *Poor*.

Finally, we plot transition probabilities for different income groups. Figures 10, 11, and 12 display transition probabilities for an individual at the 25% and at the 75% income percentile. We call them representatives of the 1. income quartile and the 3. income quartile. Surprisingly, we find that the probabilities are very similar. Higher income groups have better probabilities to remain in good health or transition towards good health but the effects differences to the low income groups are small.

4.1 Problems with Ordered Logit

The model fails the test for the parallel regression assumption. The parallel regression assumption is also called the proportional odds assumption for the ordered logit model. It describes the implicit model feature that the slope coefficients β of a series of binary regressions that can be derived from the ordered logit model have to be equal.[7] This failure hints at a probable misspecification of the model.

In order to rule out that the misspecification is not the result of an omitted variable bias we augment the model with additional regressors like individual income, individual income squared, employment status, parental mortality, several life-style choices and medical expenses (out of pocket and total). We again have to impute values for those agents who died and did not report values of time varying covariates in the year of death. As stated earlier we make the simple but strong assumption that covariates have not changed from the last survey contact two years prior with the exception of employment status and whether work requires vigorous physical activity. We set both of these indicator variables equal to zero.

[7] The following explanation follows (Long and Freese, 2006, p. 197-200). From the ordered logit model we get the probability of observing outcome $y=m$, if $\tau_{m-1} \leq y^* < \tau_m$, where τ_m are unobserved thresholds as

$$\Pr(y=m|x) = \Pr(\tau_{m-1} \leq y^* < \tau_m|x),$$

then substituting the structural model

$$y^* = x\beta + \varepsilon,$$

for latent variable y^* we get the familiar

$$\Pr(y=m|x) = F(\tau_m - x\beta) - F(\tau_{m-1} - x\beta).$$

From this expression we can calculate the cumulative probabilities

$$\Pr(y \leq m|x) = F(\tau_m - x\beta), \text{ for } m = 1, ..., J-1,$$

which can be written as a series of binary regressions for each outcome m

$$\begin{aligned} \Pr(y \leq 1|x) &= F(\tau_1 - x\beta), \\ &\vdots \\ \Pr(y \leq J-1|x) &= F(\tau_{J-1} - x\beta). \end{aligned}$$

Note that the intercept is zero so that we can identify the model. In addition, slope coefficient β is the same for all outcomes. We can test this assumption with a likelihood ratio test (in Stata command: `omodel`) as well as with a Wald test (Stata command: `brant`).

Since the extended model still fails the test of the parallel regression assumption we also estimate the multinomial logit model for the extended set of regressors. The multinomial logit model does not make the proportional odds assumption. However, the multinomial logit model carries the IIA assumption (independence of irrelevant alternatives).

A Hausman test of IIA does not reject the IIA assumption. A Small Hsiao test of IIA does reject IIA under certain sample divisions. The Small-Hsiao test for IIA divides the sample randomly into two subsamples fitting a restricted and an unrestricted model. Depending on how the sample is subdivided the test gives different results.[8]

Table 13 and table 14 report the results for the multinomial logit model with the extended list of regressors. We see that the mean probabilities do not differ much from earlier estimates using ordered logit and probit models reported in tables 9, 10, 11, and 12.

5 Survival Analysis

Our final method uses non-parametric and semi-parametric survival analysis to calculate health state hazard rates. We also describe briefly how cumulative hazard rates (derived from various health state transitions) can be related to Markov switching probabilities following the exposition in Anderson, Hansen and Keiding (1991). We follow their exposition and define

$$N_{hji}(t) = \{\text{Number of direct } h \rightarrow j \text{ transitions observed in } [0,t] \text{ for individual } i\},$$

and the health state indicator function is

$$Y_{hi}(t) = I\{\text{individual } i \text{ was observed to be in state } h \text{ at time } t-\},$$

where I is an indicator function. Then $\tilde{Y}_1, ..., \tilde{Y}_n$ are possibly right censored survival times, $D_1, ..., D_n$ are failure indicator variables and $\lambda_{hji}(t)$ defines the hazard rate (transition intensity) of individual i from state h to state j as

$$\lambda_{hji}(t) = \lim_{\Delta t \rightarrow \infty} \frac{1}{\Delta t} \Pr\{X_i(t + \Delta t) = j | X_i(t) = h\}, \ h, j = 1, 2, 3, 4, 5, 6,$$

where $X_i(t)$ are time dependent covariates. If in addition we define $\lambda_{6ji}(t) = 0$ for all $j = 1, 2, 3, 4, 5$ then we have a Markov model with absorptive state 6. States 1 through to 5 define the respective health states *Excellent*, *Very Good*, *Good*, *Fair*, *Poor* and state 6 is *Death*. We can then express

$$\begin{aligned}
Y_{hi}(t) &= I\left\{X_i(t) = h, \tilde{Y}_i \geq t\right\}, \ h = 1, 2, 3, 4, 5 \\
Y_{6i}(t) &= I\left\{\tilde{Y}_i < t, D_i = 1\right\}, \\
N_{h6i}(t) &= I\left\{\tilde{Y}_i \leq t, D_i = 1, X_i\left(\tilde{Y}_i\right) = h\right\}, \ h = 1, 2, 3, 4, 5 \\
N_{hji}(t) &= \#\left\{T_j \leq t \text{ and } \tilde{Y}_i : X_i(T_j+) = j, X_i(T_j) = h\right\}, \ h, j = 1, 2, 3, 4, 5, \\
N_{6ji}(t) &= 0, \ j = 1, 2, 3, 4, 5, 6.
\end{aligned}$$

For a group of homogenous individuals, $\lambda_{hji} = \lambda_{hj}$ we can then define the cumulative hazard as

$$\Lambda_{hj}(t) = \int_0^t \lambda_{hj}(u)\, du.$$

It can then be shown that the transition probability matrix $P(s,t) = [P_{hj}(s,t);\ h, j = 1, ..., k]$ with elements defined as $P_{hj}(s,t) = \Pr[Y_i(t) = j | Y_i(s) = h]$, $s < h$ is given by the product integral

$$P(s,t) = \prod\nolimits_{(s,t]} (I + d\Lambda(u)),$$

[8] See (Long and Freese, 2006, p. 243-246) for details.

which can be estimated as

$$\hat{P}(s,t) = \prod\nolimits_{(s,t]} \left(I + d\hat{\Lambda}(u) \right),$$

where $\hat{\Lambda}_{hj}(u)$ is the Nelson-Aalen estimator of the cumulative hazard when $h \neq j$ and $\hat{A}_{hh}(h) = -\sum_{j \neq h} \hat{A}_{hj}(u)$. This can then be expressed as the finite matrix product

$$\hat{P}(s,t) = \prod\nolimits_{T_l \in (s,t]} \left(I + \Delta\hat{\Lambda}(T_l) \right),$$

taken over the observed times T_j of transition in $(s,t]$. The $\Delta\hat{\Lambda}(T_l)$ is the $(k \times k)$ matrix with element (h,j) equal to $Y_h(T_l)^{-1}$, with diagonal elements (h,h) equal to $-Y_h(T_l)^{-1}$ and the rest of the elements equal to zero if the transition observed at time T_l is from state h to state j.[9] We will begin the analysis by deriving hazard rates for the various health state transitions.

5.1 Nonparametric Estimation of a Competing Risk Model - Hazard Rates from the Kaplan Meier Estimator

Duration in a particular health state h is defined as t_h. We next define multiple failure possibilities, that will describe the transition from health state h to health state j, where both $h, j = 1, ..., 6$. Failure of individual i at age t is denoted $f_{hji}(t)$ and defined as follows

$$f_{hji}(t) = I(Y_i(t) = j, Y_i(t-1) = h).$$

We split the sample into 5 initial health states, $h = \{Excellent, VeryGood, Good, Fair, \text{ or } Poor\}$. Out of each one of these states, we define the failure indicator $f_{hji}(t)$ for individual i with initial health state h. We then estimate the hazard rate $\lambda_{hj}(t)$ using a kernel smoother on the non-parametric estimate of the cumulative hazard rate (the Nelson-Aalen estimator). We report the resulting hazard rates between various health states in figures 13, 14, 15, 16, and 17.

Generally we can observe that women have somewhat better chances to stay in better health states, especially at ages below 60 (see crossing in figure 13, panel 1). From the non-parametric hazard rates it is difficult to see whether women really have significantly better health prospects than men. If individuals are initially in *Poor* health state, it seems that men have better recovery hazard rates than females (see 17, panel 1 and 2).

5.2 Semiparametric Estimation of a Competing Risk Model - Hazard Rates from the Proportional Hazard Framework

The assumption of homogeneity of individuals can be relaxed by adjusting the estimated hazard rates for differences in values of individual specific covariates. This can be done in a regression framework. The semi-parametric Cox type regression model is straightforward to estimate via partial likelihood methods. It can be estimated for time fixed and time varying covariates.[10]

We again follow the exposition in Anderson, Hansen and Keiding (1991) who relate the Cox hazard model to markov transition probabilities. An individual i with covariate X_i has the $h \rightarrow j$ transition hazard defined as

$$\lambda_{hji}(t) = \lambda_{hj0}(t) \exp\left(\beta_{hj} X_i\right),$$

where $\lambda_{hj0}(t)$ is the baseline hazard that is left unspecified. Then again the Markov probabilities can be estimated from the Nelson-Aalen cumulative hazard estimates

$$\hat{P}(s,t;X_0) = \prod\nolimits_{(s,t]} \left(I + d\hat{\Lambda}(u;;X_0) \right),$$

where X_0 are individuals basic covariates and X_i are individuals type specific covariates.

[9] (Anderson, Hansen and Keiding, 1991, p. 156 ff) derive the large sample properties of this estimator $\hat{P}$.

[10] Compare Cleves, Gould and Gutierrez (2004) and Cameron and Trivedi (2005) for a brief introduction to this model.

We use the same set of regressors as in the ordered logit model and define

$$X_i = \{x_{age},\ x_{age2},\ x_{female},\ x_{partner},\ x_{deducM12},\ x_{rbmi},\ x_{rsmoken},\ x_{hatota1000},\ x_{sqhatota1000}\}_i\,.$$

We report estimates of the Cox proportional hazard model in table 7.1, 7.1, and 7.1. We see that women have higher hazards to transition towards better health states (or stay in the same good health state) than men (top row in tables 7.1, 7.1). On the other hand, hazards for transition from *Poor* to *Excellent* tend to be lower for women. Whether an individual lives with a partner has a positive influence on all hazards regardless of the direction of the health transition. We find this hard to interpret. A similar results holds for the smoking indicator. Household income has a positive influence on hazards to transition from bad to better states, respectively a negative effect on hazards for transitioning from good to bad health states.

Hazard rates from the Cox proportional hazard model are reported in figures 18, 19, 20, 21, and 22 for men and women separately. From figure 18 we see that the hazard for women for transition from *Excellent* to *Excellent* is higher than that for men, whereas the hazard to transition from *Excellent* to *Poor* is lower for women. Similarly we see that women have better chances to recover from bad health states as their hazard to transition from, say, $Fair$ to *Excellent* of $Very\ Good$ is higher than that for men (see figure 21). However, if the initial health state is $Fair$ or *Poor* men seem to have a higher hazard to transition back to better health states (see figures 21, and 22). This could however be due to data limitations, since the population that is initially in bad health states is small. This confirms the earlier discussion of the Cox regression results for the female indicator variable.

Finally, we try to test the Markov assumption (health history independence of transition probabilities) by constructing a variable z_{it} that measures the time in years since last entry into the present health state. We then test whether this waiting time has any influence on the transition hazard rate. If the Markov assumption holds then the coefficient of z_{it} should not be significant.[11]

The test is unfortunately not conclusive. The optimization algorithm only converges for transitions where the initial state and the target state are identical. In such cases the Markov property does not hold. We think that there might be something wrong with the test setup. This will need further analysis.

6 Conclusion

We use the Rand-HRS dataset to estimate Markov type transition probability matrices between health states. We condition these probabilities on various agent characteristics like age, gender, education and income. We use three different methods to estimate these Markov matrices. We find that a simple count method and a regression specification using an ordered logit model produce similar results. Results from survival analysis take censoring into account and confirm the findings from the earlier analysis. Women have better prospects of staying in good health over the entire age range from 40 to 90. If initial health states are very bad, then men seem to have better recovery rates than women. Income has a positive influence on health and so does education. Smoking and body mass index have negative effects on the probabilities to either stay in good health or transition from bad to good health. The estimated Markov transition probabilities matrices can be used in life-cycle models with health uncertainty.

6.1 Informal Discussion of Issues/Problems for Further Research

- The survival analysis is still incomplete. The plotted hazard rates are not directly comparable to the transition probabilities estimated with the ordered logit approach.
- Also, I'm not sure whether the survival analysis is correct. Some of the hazard rates look puzzling. e.g. figure 17, first panel.

[11] Anderson, Hansen and Keiding (1991) propose a similar test in their liver desease study.

- The Markov test is troublesome and basically doesn't work. The algorithm doesn't converge half the time.
- Add the linear probability model and compare with results from ordered logit.
- French (2003) uses a difference approach in his linear probability model. This instruments for state dependence when he estimates the Markov probabilities and might be another method to try out.
- The parallel regression assumption is always violated (ologit - problem), which hints at a misspecified model. I wonder whether this simply comes from omitted variables. However, an even larger model with many more covariates was also rejected. I wonder whether it would be better to switch to the multinomial logit framework. There IIA seems to be less of a problem, although the Small-Hsiao test sometimes rejects IIA.
- The effects of income are far smaller than I expected. Figures 10, 11, and 12. The confidence regions overlap, so I don't think they are significantly different.

References

Anderson, Kragh, Lars S. Hansen and Niels Keiding. 1991. "Non- and Semi-Parametric Estimation of Transition Probabilities from CensoredObservationof a Non-Homogeneous Markov Process." *Scandinavian Journal of Statistics* 18:153–167.

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.

Cook, Richard J., Jerald F. Lawless and Ker-Ai Lee. 2003. "Cumulative processes related to event histories." *SORT - Statistics and Operations Research Transactions* 27:13–30.

Dabrowska, Dorota M., Guo-Wen Sun and Mary M. Horowitz. 1994. "Cox Regression in a Markov Renewal Model: An Application to the Analysis of Bone Marrow Transplant Data." *Nournal of theAmerican Statistical Association* 89(427):867–877.

De Nardi, Mariachristina, Eric French and John B. Jones. 2006. "Differential Mortality, Uncertain Medical Expenses and the Saving of theElderlySingles." Working Paper 2005-13, Federal Reserve Bank of Chicago.

Diehr, Paula and Donald L. Patrick. 2001. "Probabilities of Transition Among Health States for Older Adults." *Quality of Life Research* 10:431–442.

Diehr, Paula, Donald L. Patrick, Diane E. Bild and Gregory L. BurkeandJeff D. Williamson. 1998. "Predicting Future Years of Healthy Life for Older Adults." *Journal of Clinical Epidemiology* 51(4):343–353.

French, Eric. 2003. "The Effects of Health, Wealth, and Wages on Labor Supply and Retirement-Behavior." Federal Reserve Chicago, Working Papers Series WB 2000-02.

French, Eric and John B. Jones. 2004. "The Effects of Health INsurance and Self-Insurance on Retirement Behavior." Federal Reserve Bank of Chicago, Working Paper 2001-19.

Juster, F. Thomas and Richard Suzman. 1995. "An Overview of the Health and Retirement Study." *Journal of Human Resources* 30(Supplement):S7–S56.

Long, Scott J. and Jeremy Freese. 2006. *Regression Models for Categorical Dependent Variables Using Stata*. 2 ed. Stata Press.

Mundahl, J. M. 1998. "Imputation of Missing Longitudinal Data: A Comparison of Methods." Master's Thesis, University of Washington.

Palumbo, Michael G. 1999. "Uncertain Medical Expenses and Precautionary Saving Near the End of theLifeCycle." *Review of Economic Studies* 66(2):395–421.

Wallace, Robert B. and Regula A. Herzog. 1995. "Overview of the Health Measures in the Health and Retirement Study." *Journal of Human Resources* 30(Supplement):S84–S107.

7 Appendix:

7.1 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

Table 2: Summary Statistics of Covariates

Variable	**Mean**	**Std. Dev.**	**N**
age	66.067	11.589	135903
age2	4499.12	1585.557	135903
female	0.576	0.494	135905
partner	0.67	0.47	135776
deducM12	0.358	0.48	135560
rbmi	26.932	5.307	127827
rsmoken	0.181	0.385	122127
hatota1000	297.262	923.731	135905
sqhatota1000	941636.502	43152078.009	135905

Table 3: **Total Number of Health States per Age Group**

	Health						
age	Excellent	VeryGood	Good	Fair	Poor	Dead	Total
40	24	24	26	15	3	0	92
41	18	33	27	10	4	1	93
42	26	42	26	19	2	1	116
43	24	49	42	14	11	1	141
44	37	62	46	25	5	0	175
45	44	85	68	19	5	2	223
46	67	87	89	33	12	2	290
47	77	115	89	46	21	3	351
48	89	154	117	63	19	2	444
49	106	179	160	50	32	7	534
50	127	218	190	91	35	5	666
51	177	283	242	93	50	6	851
52	267	460	394	190	79	13	1,403
53	460	660	546	273	125	20	2,084
54	418	793	686	337	150	36	2,420
55	569	952	849	461	179	44	3,054
56	582	1,013	964	502	236	41	3,338
57	661	1,167	1,108	599	287	60	3,882
58	582	1,163	1,105	625	271	65	3,811
59	622	1,216	1,179	687	325	78	4,107
60	594	1,178	1,238	647	311	89	4,057
61	612	1,285	1,230	744	356	101	4,328
62	567	1,246	1,242	713	349	119	4,236
63	539	1,182	1,232	745	323	113	4,134
64	451	1,045	1,149	675	300	106	3,726
65	467	981	1,105	627	257	105	3,542
66	347	808	988	567	244	100	3,054
67	295	818	841	492	223	115	2,784
68	265	707	720	441	213	102	2,448
69	236	675	701	429	201	105	2,347
70	235	563	662	401	195	108	2,164
71	211	562	638	407	179	94	2,091
72	212	631	739	449	204	112	2,347
73	199	580	692	422	215	130	2,238
74	224	641	727	511	230	176	2,509
75	210	592	721	510	230	174	2,437
76	178	552	734	516	226	191	2,397
77	181	491	735	471	254	195	2,327
78	154	474	630	496	249	207	2,210
79	125	445	558	456	256	220	2,060
80	110	379	570	430	237	214	1,940
81	121	339	530	426	218	232	1,866
82	107	316	485	396	212	224	1,740
83	104	264	404	390	225	228	1,615
84	86	265	383	336	200	219	1,489
85	63	211	331	320	188	230	1,343
86	58	207	278	264	180	204	1,191
87	53	147	232	239	163	234	1,068
88	42	124	227	201	140	201	935
89	28	105	167	195	112	206	813
90	37	96	151	154	86	189	713
Total	12,088	26,664	28,993	18,222	8,827	5,430	100,224

Source: Rand-HRS 2004

Table 4: **Number of Health State Transitions per Age and Gender**

	Gender		
age	male	female	Total
40	12	80	92
41	8	85	93
42	12	104	116
43	19	122	141
44	18	157	175
45	26	197	223
46	42	248	290
47	46	305	351
48	49	395	444
49	64	470	534
50	81	585	666
51	118	733	851
52	365	1,038	1,403
53	727	1,357	2,084
54	881	1,539	2,420
55	1,215	1,839	3,054
56	1,391	1,947	3,338
57	1,664	2,218	3,882
58	1,673	2,138	3,811
59	1,827	2,280	4,107
60	1,815	2,242	4,057
61	1,934	2,394	4,328
62	1,936	2,300	4,236
63	1,925	2,209	4,134
64	1,803	1,923	3,726
65	1,698	1,844	3,542
66	1,550	1,504	3,054
67	1,362	1,422	2,784
68	1,268	1,180	2,448
69	1,140	1,207	2,347
70	1,047	1,117	2,164
71	942	1,149	2,091
72	1,098	1,249	2,347
73	1,036	1,202	2,238
74	1,158	1,351	2,509
75	1,095	1,342	2,437
76	1,093	1,304	2,397
77	1,008	1,319	2,327
78	953	1,257	2,210
79	870	1,190	2,060
80	785	1,155	1,940
81	753	1,113	1,866
82	685	1,055	1,740
83	619	996	1,615
84	545	944	1,489
85	511	832	1,343
86	411	780	1,191
87	380	688	1,068
88	319	616	935
89	285	528	813
90	229	484	713
Total	42,491	57,733	100,224

Source: Rand-HRS 2004

Table 5: **Count of Transition Probabilities: Females, Age: 40-90**

	Health (in %)						
Lag Health	Excellent	VeryGood	Good	Fair	Poor	Dead	Total
Excellent	47.3	33.1	13.0	3.6	1.1	1.9	100.0
VeryGood	13.0	48.7	27.4	6.9	1.5	2.4	100.0
Good	4.3	20.6	48.2	18.1	4.0	4.7	100.0
Fair	1.5	6.7	24.3	42.6	14.4	10.5	100.0
Poor	0.7	1.9	7.2	22.4	43.8	24.0	100.0
Total	12.2	25.8	29.2	17.8	8.5	6.5	100.0

	Health						
Lag Health	Excellent	VeryGood	Good	Fair	Poor	Dead	Total
Excellent	2,971	2,076	816	225	68	119	6,275
VeryGood	1,502	5,618	3,163	799	171	274	11,527
Good	557	2,681	6,265	2,359	525	616	13,003
Fair	117	521	1,886	3,299	1,113	812	7,748
Poor	29	75	284	881	1,723	946	3,938
Total Frequency	5,176	10,971	12,414	7,563	3,600	2,767	42,491

Table 6: **Count of Transition Probabilities: Males, Age: 40-90**

	Health (in %)						
Lag Health	Excellent	VeryGood	Good	Fair	Poor	Dead	Total
Excellent	50.5	33.9	11.2	2.4	0.7	1.2	100.0
VeryGood	12.7	51.7	26.8	5.7	1.5	1.6	100.0
Good	3.4	21.8	49.1	18.4	4.0	3.2	100.0
Fair	1.3	6.4	23.5	46.1	15.4	7.2	100.0
Poor	0.4	1.8	7.2	25.9	47.0	17.7	100.0
Total	12.0	27.2	28.7	18.5	9.1	4.6	100.0

	Health						
Lag Health	Excellent	VeryGood	Good	Fair	Poor	Dead	Total
Excellent	4,099	2,756	911	195	60	98	8,119
VeryGood	2,063	8,431	4,372	934	243	265	16,308
Good	585	3,717	8,358	3,133	689	552	17,034
Fair	145	688	2,542	4,982	1,662	780	10,799
Poor	20	101	396	1,415	2,573	968	5,473
Total Frequency	6,912	15,693	16,579	10,659	5,227	2,663	57,733

Table 7: Transition Probabilities from Ordered Logit Model, Females (Age: 40-90)

	Final health	1 Excellent	2 Very Good	3 Good	4 Fair	5 Poor	Sum	N
Initial health	.	.	.	.	.	.	.	.
1 Excellent	.	0.5142	0.3317	0.1168	0.0287	0.0086	1.0000	7314.0000
2 Very Good	.	0.1361	0.5120	0.2736	0.0632	0.0152	1.0000	1.42e+04
3 Good	.	0.0403	0.2218	0.5037	0.1917	0.0425	1.0000	1.45e+04
4 Fair	.	0.0154	0.0701	0.2548	0.4908	0.1689	1.0000	8926.0000
5 Poor	.	0.0061	0.0231	0.0887	0.3043	0.5777	1.0000	4071.0000

Table 8: Transition Probabilities from Ordered Logit Model, Males (Age: 40-90)

	Final health	1 Excellent	2 Very Good	3 Good	4 Fair	5 Poor	Sum	N
Initial health	.	.	.	.	.	.	.	.
1 Excellent	.	0.4713	0.3534	0.1325	0.0330	0.0099	1.0000	5807.0000
2 Very Good	.	0.1270	0.5073	0.2837	0.0661	0.0159	1.0000	1.05e+04
3 Good	.	0.0402	0.2213	0.5049	0.1914	0.0422	1.0000	1.17e+04
4 Fair	.	0.0167	0.0758	0.2676	0.4832	0.1567	1.0000	6561.0000
5 Poor	.	0.0063	0.0239	0.0912	0.3086	0.5700	1.0000	2852.0000

Table 9: Transition Probabilities from Ordered Logit Model including Death, Females (Age: 40-90)

	Final health	1 Excellent	2 Very Good	3 Good	4 Fair	5 Poor	6 Dead	Sum	N
Initial health	.	.	.	.	.	.	.	.	.
1 Excellent	.	0.4603	0.3477	0.1316	0.0329	0.0099	0.0176	1.0000	5924.0000
2 Very Good	.	0.1225	0.4932	0.2803	0.0663	0.0160	0.0218	1.0000	1.08e+04
3 Good	.	0.0373	0.2065	0.4821	0.1885	0.0422	0.0433	1.0000	1.23e+04
4 Fair	.	0.0142	0.0644	0.2316	0.4435	0.1530	0.0934	1.0000	7359.0000
5 Poor	.	0.0044	0.0165	0.0631	0.2211	0.4622	0.2327	1.0000	3781.0000

Table 10: Transition Probabilities from Ordered Logit Model including Death, Males (Age: 40-90)

	Final health	1 Excellent	2 Very Good	3 Good	4 Fair	5 Poor	6 Dead	Sum	N
Initial health	.	.	.	.	.	.	.	.	.
1 Excellent	.	0.5075	0.3255	0.1150	0.0283	0.0085	0.0151	1.0000	7411.0000
2 Very Good	.	0.1344	0.5007	0.2672	0.0623	0.0150	0.0204	1.0000	1.45e+04
3 Good	.	0.0396	0.2149	0.4801	0.1828	0.0408	0.0418	1.0000	1.50e+04
4 Fair	.	0.0147	0.0667	0.2361	0.4405	0.1502	0.0918	1.0000	9683.0000
5 Poor	.	0.0053	0.0199	0.0744	0.2434	0.4505	0.2064	1.0000	5013.0000

Table 11: Transition Probabilities from Ordered Probit Model including Death, Females (Age: 40-90)

	Final health	1 Excellent	2 Very Good	3 Good	4 Fair	5 Poor	6 Dead	Sum	N
Initial health	.	.	.	.	.	.	.	.	.
1 Excellent	.	0.4617	0.3467	0.1313	0.0330	0.0099	0.0174	1.0000	5924.0000
2 Very Good	.	0.1234	0.4951	0.2792	0.0657	0.0158	0.0207	1.0000	1.08e+04
3 Good	.	0.0378	0.2076	0.4828	0.1878	0.0421	0.0419	1.0000	1.23e+04
4 Fair	.	0.0144	0.0642	0.2313	0.4442	0.1534	0.0926	1.0000	7359.0000
5 Poor	.	0.0046	0.0171	0.0643	0.2254	0.4611	0.2275	1.0000	3781.0000

Table 12: Transition Probabilities from Ordered Probit Model including Death, Males (Age: 40-90)

	Final health	1 Excellent	2 Very Good	3 Good	4 Fair	5 Poor	6 Dead	Sum	N
Initial health	.	.	.	.	.	.	.	.	.
1 Excellent	.	0.5112	0.3264	0.1134	0.0272	0.0080	0.0138	1.0000	7411.0000
2 Very Good	.	0.1362	0.5010	0.2676	0.0614	0.0146	0.0191	1.0000	1.45e+04
3 Good	.	0.0410	0.2152	0.4812	0.1823	0.0404	0.0400	1.0000	1.50e+04
4 Fair	.	0.0158	0.0681	0.2373	0.4412	0.1489	0.0887	1.0000	9683.0000
5 Poor	.	0.0064	0.0215	0.0753	0.2428	0.4518	0.2022	1.0000	5013.0000

Table 13: Transition Probabilities from a Multinomial Logit Model including Death, Females (Age: 40-90)

	Final health	1 Excellent	2 Very Good	3 Good	4 Fair	5 Poor	6 Dead	Sum	N
Initial health	.	.	.	.	.	.	.	.	.
1 Excellent	.	0.4764	0.3229	0.1322	0.0357	0.0111	0.0217	1.0000	5209.0000
2 Very Good	.	0.1160	0.4210	0.2403	0.1683	0.0273	0.0270	1.0000	9231.0000
3 Good	.	0.0455	0.2076	0.4726	0.1768	0.0403	0.0573	1.0000	1.04e+04
4 Fair	.	0.0152	0.0646	0.2450	0.4111	0.1383	0.1258	1.0000	6271.0000
5 Poor	.	0.0078	0.0182	0.0705	0.2181	0.4146	0.2709	1.0000	3363.0000

Table 14: Transition Probabilities from a Multinomial Logit Model including Death, Males (Age: 40-90)

	Final health	1 Excellent	2 Very Good	3 Good	4 Fair	5 Poor	6 Dead	Sum	N
Initial health	.	.	.	.	.	.	.	.	.
1 Excellent	.	0.5047	0.3385	0.1127	0.0234	0.0069	0.0139	1.0000	6834.0000
2 Very Good	.	0.1135	0.4607	0.2397	0.1444	0.0231	0.0186	1.0000	1.30e+04
3 Good	.	0.0353	0.2205	0.4843	0.1798	0.0406	0.0396	1.0000	1.35e+04
4 Fair	.	0.0145	0.0663	0.2332	0.4476	0.1515	0.0869	1.0000	8690.0000
5 Poor	.	0.0024	0.0178	0.0702	0.2472	0.4579	0.2044	1.0000	4603.0000

Initial State →	(1) Excellent	(2) VGood	(3) Good	(4) Fair	(5) Poor
1 P(Exc|.)					
age	0.0300***	0.0160***	0.00578***	0.00147***	0.000226
	(0.0059)	(0.0023)	(0.00075)	(0.00037)	(0.00015)
age2	-0.000290***	-0.000163***	-0.0000569***	-0.0000153***	-0.00000301***
	(0.000045)	(0.000017)	(0.0000055)	(0.0000027)	(0.0000011)
female	0.0338***	0.0177***	0.00557***	0.00167**	0.000910***
	(0.010)	(0.0044)	(0.0015)	(0.00073)	(0.00032)
partner	0.0142	0.0142***	0.00437***	0.00199***	-0.000422
	(0.012)	(0.0048)	(0.0015)	(0.00070)	(0.00030)
deducM12	0.116***	0.0442***	0.0130***	0.00337***	-0.000627*
	(0.010)	(0.0038)	(0.0015)	(0.00083)	(0.00038)
rbmi	-0.0128***	-0.00531***	-0.000925***	0.0000321	0.000125***
	(0.0014)	(0.00050)	(0.00015)	(0.000058)	(0.000029)
rsmoken	-0.0978***	-0.0436***	-0.0153***	-0.00305***	-0.000384
	(0.014)	(0.0051)	(0.0017)	(0.00083)	(0.00031)
hatota1000	0.0000599***	0.0000270***	0.00000770***	0.00000613***	0.00000123*
	(0.0000084)	(0.0000043)	(0.0000015)	(0.0000014)	(0.00000071)
sqhatota1000	-1.49e-09***	-3.68e-10**	-9.60e-11**	-3.37e-10**	-1.99e-10**
	(2.61e-10)	(1.75e-10)	(0)	(1.52e-10)	(9.05e-11)
2 P(VGood|.)					
age	-0.0177***	0.00315***	0.0174***	0.00565***	0.000823
	(0.0033)	(0.0011)	(0.0025)	(0.0014)	(0.00053)
age2	0.000171***	-0.0000322***	-0.000172***	-0.0000587***	-0.0000110***
	(0.000025)	(0.000010)	(0.000019)	(0.000010)	(0.0000037)
female	-0.0205***	0.00224**	0.0162***	0.00634**	0.00330***
	(0.0061)	(0.00099)	(0.0044)	(0.0028)	(0.0011)
partner	-0.00828	0.00373**	0.0136***	0.00774***	-0.00153
	(0.0070)	(0.0017)	(0.0048)	(0.0027)	(0.0011)
deducM12	-0.0607***	0.0186***	0.0429***	0.0132***	-0.00228*
	(0.0055)	(0.0028)	(0.0045)	(0.0031)	(0.0013)
rbmi	0.00759***	-0.00105***	-0.00279***	0.000123	0.000455***
	(0.00090)	(0.00025)	(0.00041)	(0.00022)	(0.000083)
rsmoken	0.0524***	-0.0183***	-0.0514***	-0.0119***	-0.00140
	(0.0074)	(0.0034)	(0.0054)	(0.0032)	(0.0011)
hatota1000	-0.0000354***	0.00000534***	0.0000232***	0.0000235***	0.00000448*
	(0.0000048)	(0.0000017)	(0.0000046)	(0.0000051)	(0.0000025)
sqhatota1000	8.80e-10***	-7.28e-11*	-2.90e-10**	-1.29e-09**	-7.25e-10**
	(1.52e-10)	(0)	(1.23e-10)	(5.80e-10)	(3.11e-10)
Observations	13332	25245	27308	17033	8788

Standard errors in parentheses

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

Table 15: Marginal Effects from Ordered Logit: Age: 40-90

Initial State →	(1) Excellent	(2) VGood	(3) Good	(4) Fair	(5) Poor
3 P(Good|.)					
age	-0.00849***	-0.0138***	-0.00898***	0.00932***	0.00274
	(0.0018)	(0.0022)	(0.0011)	(0.0026)	(0.0018)
age2	0.0000820***	0.000141***	0.0000884***	-0.0000968***	-0.0000365***
	(0.000014)	(0.000017)	(0.0000082)	(0.000020)	(0.000012)
female	-0.00923***	-0.0145***	-0.00898***	0.0101**	0.0108***
	(0.0029)	(0.0037)	(0.0024)	(0.0044)	(0.0035)
partner	-0.00407	-0.0129***	-0.00653***	0.0134***	-0.00506
	(0.0035)	(0.0045)	(0.0022)	(0.0048)	(0.0034)
deducM12	-0.0374***	-0.0442***	-0.0172***	0.0236***	-0.00749*
	(0.0038)	(0.0039)	(0.0025)	(0.0053)	(0.0044)
rbmi	0.00363***	0.00460***	0.00144***	0.000203	0.00151***
	(0.00041)	(0.00039)	(0.00025)	(0.00037)	(0.00025)
rsmoken	0.0309***	0.0436***	0.0193***	-0.0211***	-0.00469
	(0.0050)	(0.0053)	(0.0028)	(0.0056)	(0.0036)
hatota1000	-0.0000170***	-0.0000234***	-0.0000120***	0.0000388***	0.0000149*
	(0.0000026)	(0.0000038)	(0.0000024)	(0.0000089)	(0.0000082)
sqhatota1000	4.21e-10***	3.19e-10**	1.49e-10**	-2.14e-09**	-2.41e-09**
	(7.75e-11)	(1.53e-10)	(6.35e-11)	(9.78e-10)	(1.03e-09)
4 P(Fair|.)					
age	-0.00207***	-0.00339***	-0.00973***	-0.00745***	0.00514
	(0.00046)	(0.00059)	(0.0016)	(0.0018)	(0.0036)
age2	0.0000200***	0.0000347***	0.0000958***	0.0000775***	-0.0000685***
	(0.0000037)	(0.0000047)	(0.000012)	(0.000013)	(0.000026)
female	-0.00223***	-0.00348***	-0.00876***	-0.00849**	0.0190***
	(0.00070)	(0.00090)	(0.0024)	(0.0037)	(0.0060)
partner	-0.00100	-0.00324***	-0.00779***	-0.00997***	-0.00921
	(0)	(0.0012)	(0.0028)	(0.0034)	(0.0061)
deducM12	-0.00967***	-0.0118***	-0.0261***	-0.0166***	-0.0134*
	(0.0011)	(0.0012)	(0.0027)	(0.0041)	(0.0081)
rbmi	0.000888***	0.00113***	0.00156***	-0.000163	0.00284***
	(0.00011)	(0.000098)	(0.00023)	(0.00030)	(0.00046)
rsmoken	0.00791***	0.0117***	0.0318***	0.0151***	-0.00907
	(0.0014)	(0.0015)	(0.0035)	(0.0041)	(0.0070)
hatota1000	-0.00000414***	-0.00000574***	-0.0000130***	-0.0000310***	0.0000280*
	(0.00000068)	(0.00000097)	(0.0000026)	(0.0000068)	(0.000016)
sqhatota1000	1.03e-10***	7.83e-11**	1.62e-10**	1.71e-09**	-4.53e-09**
	(0)	(0)	(6.94e-11)	(7.64e-10)	(1.96e-09)
Observations	13332	25245	27308	17033	8788

Standard errors in parentheses

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

Table 16: Marginal Effects from Ordered Logit Continued: Age: 40-90

Initial State →	(1) Excellent	(2) VGood	(3) Good	(4) Fair	(5) Poor
5 P(Poor|.)					
age	-0.000617***	-0.000805***	-0.00223***	-0.00548***	-0.00376
	(0.00015)	(0.00015)	(0.00038)	(0.0016)	(0.0023)
age2	0.00000597***	0.00000822***	0.0000219***	0.0000570***	0.0000501***
	(0.0000012)	(0.0000012)	(0.0000030)	(0.000012)	(0.000016)
female	-0.000661***	-0.000821***	-0.00198***	-0.00587**	-0.0155***
	(0.00022)	(0.00021)	(0.00055)	(0.0026)	(0.0051)
partner	-0.000298	-0.000773***	-0.00180***	-0.00798***	0.00711
	(0.00026)	(0.00028)	(0.00067)	(0.0029)	(0.0050)
deducM12	-0.00292***	-0.00286***	-0.00624***	-0.0142***	0.0106*
	(0.00041)	(0.00031)	(0.00068)	(0.0032)	(0.0062)
rbmi	0.000264***	0.000268***	0.000356***	-0.000120	-0.00208***
	(0.000037)	(0.000026)	(0.000052)	(0.00022)	(0.00040)
rsmoken	0.00238***	0.00281***	0.00768***	0.0127***	0.00627
	(0.00045)	(0.00039)	(0.00091)	(0.0034)	(0.0049)
hatota1000	-0.00000123***	-0.00000136***	-0.00000297***	-0.0000228***	-0.0000205*
	(0.00000022)	(0.00000024)	(0.00000062)	(0.0000053)	(0.000011)
sqhatota1000	0***	0**	0**	1.26e-09**	3.31e-09**
	(0)	(0)	(0)	(5.79e-10)	(1.42e-09)
6 P(Dead|.)					
age	-0.00108***	-0.00108***	-0.00226***	-0.00350***	-0.00517
	(0.00025)	(0.00020)	(0.00040)	(0.0011)	(0.0037)
age2	0.0000105***	0.0000111***	0.0000223***	0.0000364***	0.0000689**
	(0.0000021)	(0.0000016)	(0.0000031)	(0.0000082)	(0.000028)
female	-0.00116***	-0.00110***	-0.00200***	-0.00371**	-0.0185***
	(0.00037)	(0.00029)	(0.00056)	(0.0017)	(0.0060)
partner	-0.000523	-0.00104***	-0.00184***	-0.00518***	0.00912
	(0.00046)	(0.00038)	(0.00069)	(0.0019)	(0.0060)
deducM12	-0.00516***	-0.00387***	-0.00643***	-0.00936***	0.0132
	(0.00064)	(0.00040)	(0.00070)	(0.0021)	(0.0081)
rbmi	0.000463***	0.000360***	0.000362***	-0.0000764	-0.00286***
	(0.000058)	(0.000034)	(0.000053)	(0.00014)	(0.00050)
rsmoken	0.00420***	0.00381***	0.00794***	0.00831***	0.00928
	(0.00077)	(0.00053)	(0.00096)	(0.0023)	(0.0072)
hatota1000	-0.00000216***	-0.00000183***	-0.00000302***	-0.0000146***	-0.0000282*
	(0.00000037)	(0.00000032)	(0.00000063)	(0.0000035)	(0.000016)
sqhatota1000	5.37e-11***	0**	0**	8.03e-10**	4.56e-09**
	(0)	(0)	(0)	(3.73e-10)	(2.01e-09)
Observations	13332	25245	27308	17033	8788

Standard errors in parentheses

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

Table 17: Marginal Effects from Ordered Logit Continued: Age: 40-90

COEFFICIENT	(1) Exc → Exc	(2) Exc → Exc	(3) Exc → Good	(4) Exc → Good	(5) Exc → Poor	(6) Exc → Poor
female	0.340***	0.166***	0.0552	0.106**	-0.450**	-0.428**
	(0.026)	(0.027)	(0.052)	(0.053)	(0.20)	(0.20)
partner	0.698***	0.465***	0.472***	0.530***	0.136	0.181
	(0.032)	(0.032)	(0.060)	(0.060)	(0.21)	(0.21)
deducM12	0.529***	0.247***	-0.215***	-0.0380	-0.168	0.0114
	(0.026)	(0.027)	(0.054)	(0.054)	(0.20)	(0.20)
rbmi	-0.00984***	-0.00228	0.0523***	0.0519***	0.0414**	0.0376*
	(0.0030)	(0.0030)	(0.0043)	(0.0052)	(0.019)	(0.021)
rsmoken	0.198***	0.186***	0.615***	0.580***	0.463*	0.468*
	(0.038)	(0.038)	(0.069)	(0.069)	(0.28)	(0.28)
hatota1000	0.0000885***	-0.0000627***	-0.000175***	-0.0000761	-0.00168***	-0.00154***
	(0.000017)	(0.000017)	(0.000054)	(0.000052)	(0.00044)	(0.00043)
sqhatota1000	-2.30e-09***	1.79e-09***	3.81e-09*	1.14e-09	0.0000000393**	0.0000000358**
	(7.15e-10)	(6.37e-10)	(2.09e-09)	(2.18e-09)	(0.000000016)	(0.000000016)
jmarkov		0.364***		-26.37		-24.77
		(0.0035)		(0)		(0)
Observations	13335	13335	13335	13335	13335	13335
PropHaz Chi2 Test	174.0		51.85		12.79	

Standard errors in parentheses

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

Table 18: Cox Proportional Hazard Model. Transition from health state Excellent to various health states. Age: 40-90

COEFFICIENT	(1) Good → Exc	(2) Good → Exc	(3) Good → Good	(4) Good → Good	(5) Good → Poor	(6) Good → Poor
female	-0.0188	0.0207	0.224***	0.165***	0.0750	0.113*
	(0.065)	(0.065)	(0.018)	(0.018)	(0.065)	(0.065)
partner	0.503***	0.571***	0.648***	0.477***	0.488***	0.545***
	(0.073)	(0.073)	(0.021)	(0.021)	(0.069)	(0.069)
deducM12	0.319***	0.345***	0.201***	0.112***	-0.244***	-0.203***
	(0.065)	(0.066)	(0.019)	(0.019)	(0.074)	(0.074)
rbmi	0.0246***	0.0331***	0.0600***	0.0394***	0.0284***	0.0350***
	(0.0063)	(0.0061)	(0.0016)	(0.0017)	(0.0061)	(0.0060)
rsmoken	0.499***	0.540***	0.670***	0.454***	0.799***	0.830***
	(0.083)	(0.083)	(0.023)	(0.023)	(0.080)	(0.080)
hatota1000	-0.0000208	-0.0000144	-0.0000595***	-0.0000797***	-0.000682***	-0.000631***
	(0.000054)	(0.000060)	(0.000016)	(0.000018)	(0.00011)	(0.00012)
sqhatota1000	-7.34e-11	-4.55e-10	7.07e-10**	1.06e-09***	7.91e-09***	7.26e-09***
	(2.01e-09)	(3.67e-09)	(2.98e-10)	(2.88e-10)	(1.47e-09)	(1.66e-09)
jmarkov		-26.12		0.411***		-20.04
		(0)		(0.0026)		(9950111)
Observations	27311	27311	27311	27311	27311	27311
PropHaz Chi2 Test	31.57		160.1		63.68	

Standard errors in parentheses

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

Table 19: Cox Proportional Hazard Model. Transition from health state Good to various health states. Age: 40-90

COEFFICIENT	(1) Poor → Exc	(2) Poor → Exc	(3) Poor → Good	(4) Poor → Good	(5) Poor → Poor	(6) Poor → Po
female	-0.829**	-0.769**	0.0970	0.160*	0.193***	0.146***
	(0.35)	(0.34)	(0.088)	(0.088)	(0.035)	(0.035)
partner	0.219	0.217	0.415***	0.405***	0.405***	0.387***
	(0.34)	(0.34)	(0.091)	(0.090)	(0.035)	(0.035)
deducM12	0.153	0.125	0.175*	0.135	-0.0581	0.0222
	(0.39)	(0.39)	(0.11)	(0.11)	(0.046)	(0.046)
rbmi	0.0257	0.0324	0.0498***	0.0553***	0.0621***	0.0309***
	(0.024)	(0.023)	(0.0057)	(0.0053)	(0.0020)	(0.0023)
rsmoken	0.0849	0.145	0.708***	0.741***	0.831***	0.561***
	(0.42)	(0.42)	(0.10)	(0.10)	(0.039)	(0.040)
hatota1000	0.00000814	-0.000141	-0.000248	-0.000388*	-0.000497***	-0.000265**
	(0.00087)	(0.00080)	(0.00025)	(0.00021)	(0.000087)	(0.000075)
sqhatota1000	-0.0000000457	-0.0000000158	-0.0000000144	6.03e-09	0.0000000151***	7.70e-09**
	(0.00000028)	(0.00000020)	(0.000000079)	(0.000000029)	(3.41e-09)	(3.41e-09)
jmarkov		-18.59		-22.56		0.445***
		(12562031)		(177879445)		(0.0049)
Observations	8794	8794	8794	8794	8794	8794
PropHaz Chi2 Test	2.278		9.591		72.06	

Standard errors in parentheses

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

Table 20: Cox Proportional Hazard Model. Transition from health state Good to various health states. Age: 40-90

7.2 Figures

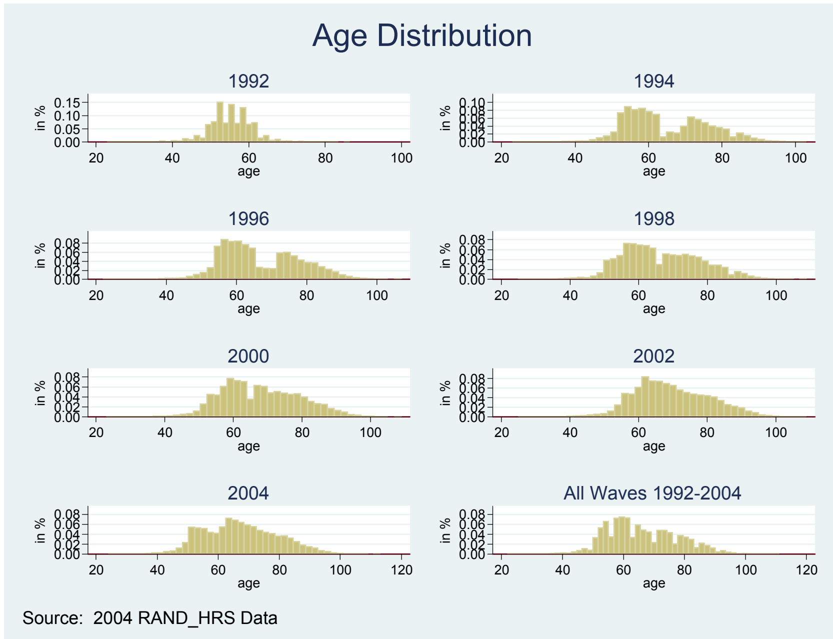

Figure 1: Histogram of Age over all Waves.

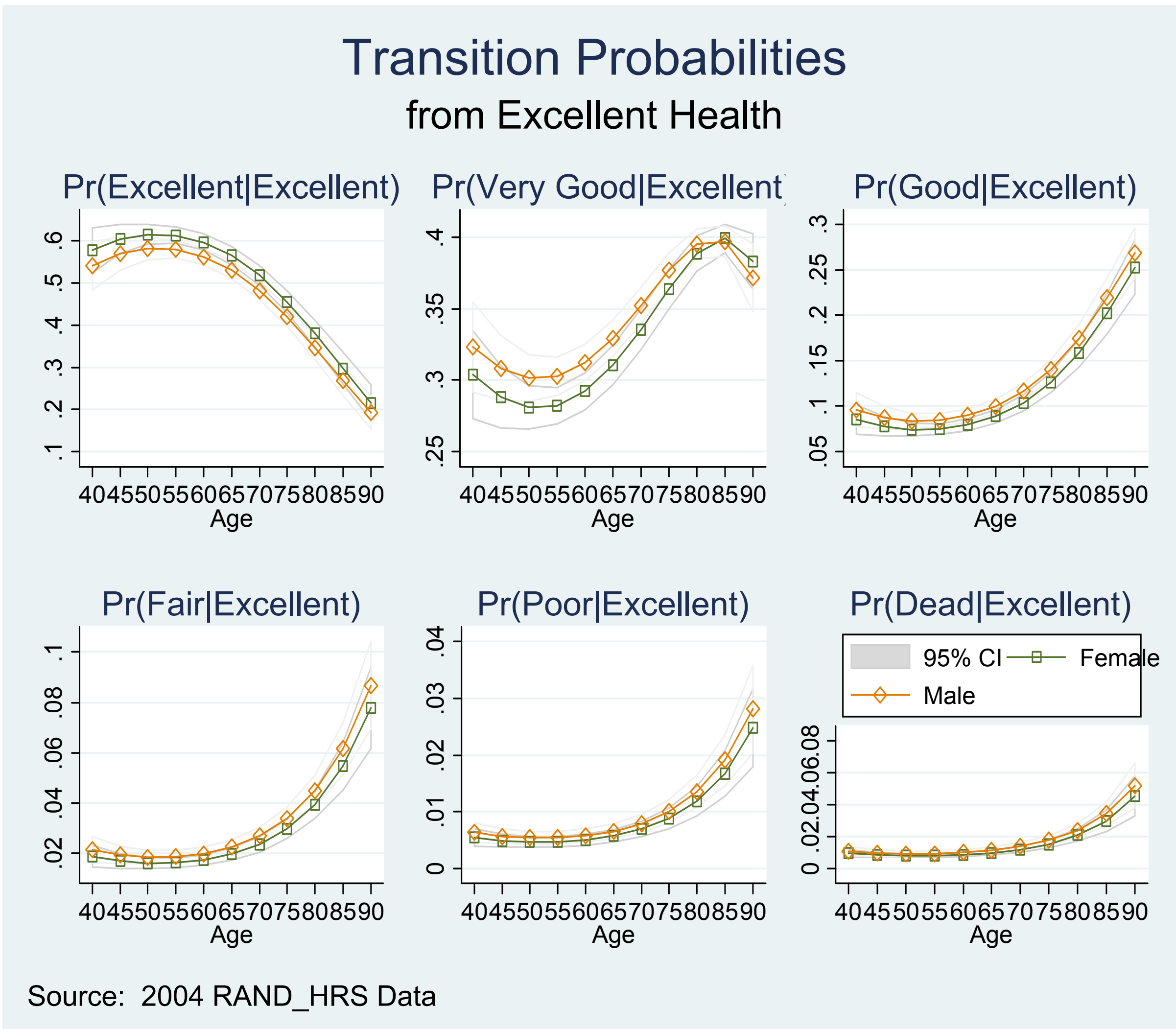

Figure 2: Transition Probabilities from Excellent Health State to Other Health States. We report the predictions at the median of household income. Predictions based on ordered Logit estimates for agegroup: 40-90.

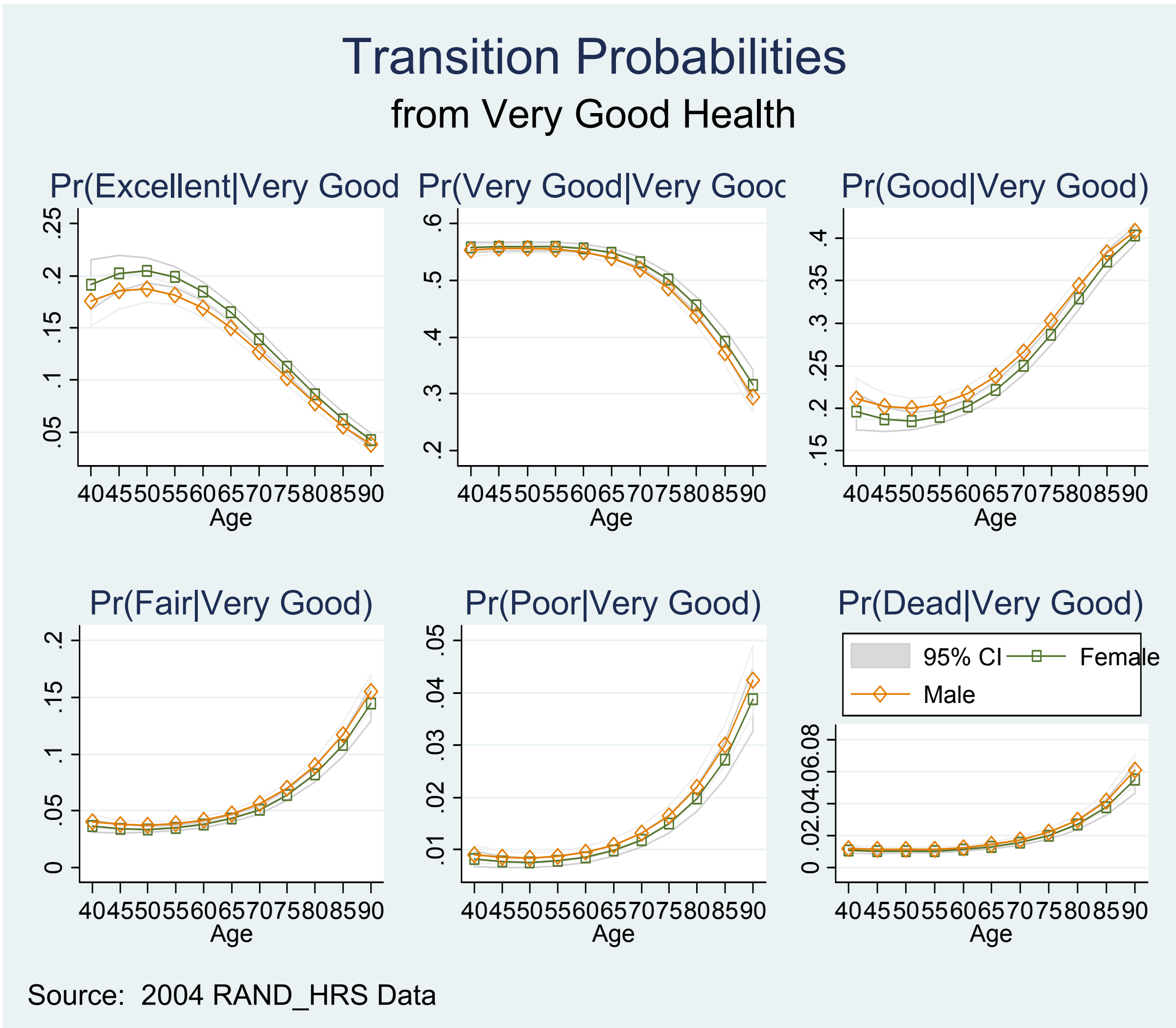

Figure 3: Transition Probabilities from Very Good Health State to Other Health States. We report the predictions at the median of household income. Predictions based on ordered Logit estimates for agegroup: 40-90.

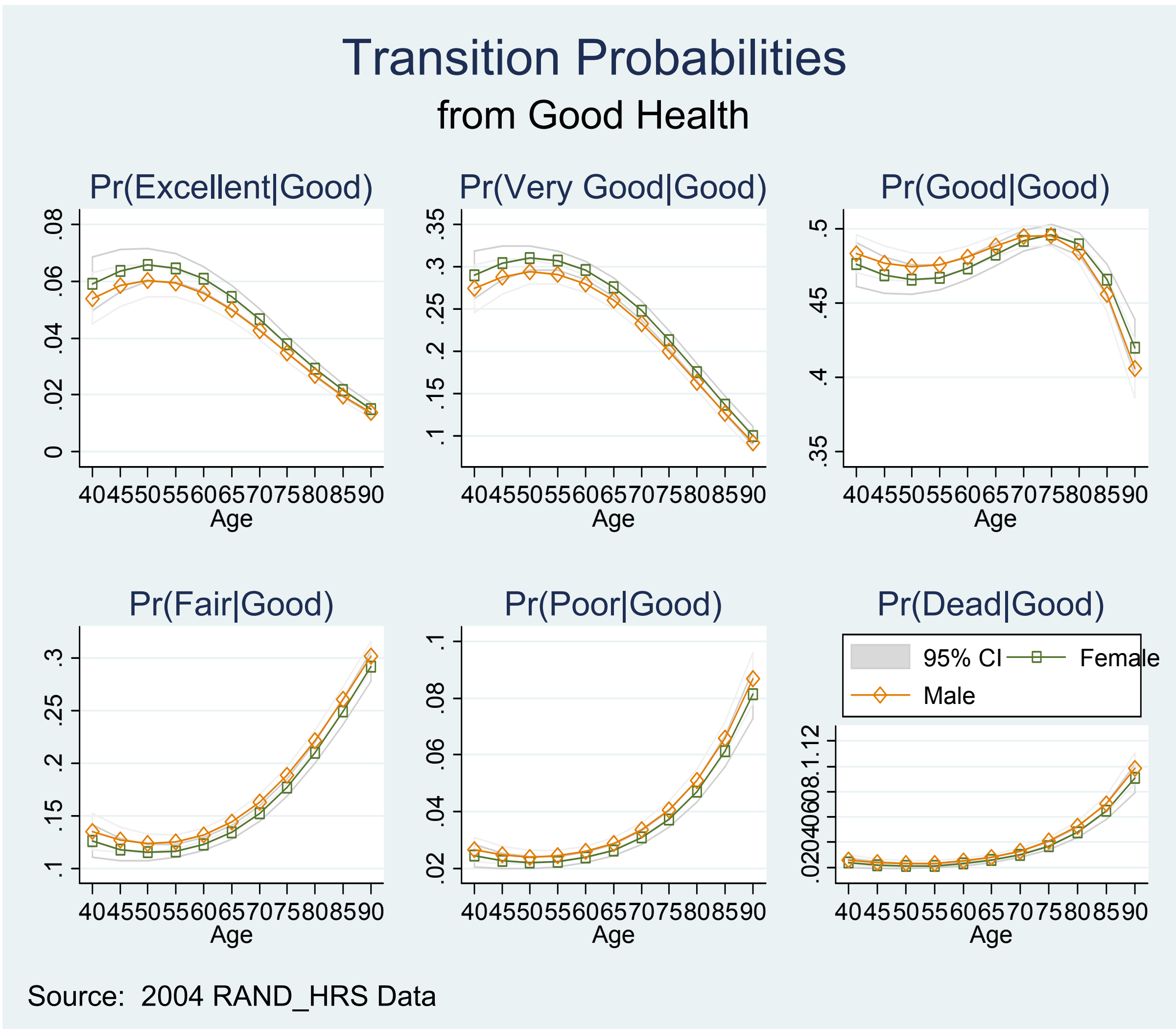

Figure 4: Transition Probabilities from Good Health State to Other Health States. We report the predictions at the median of household income. Predictions based on ordered Logit estimates for agegroup: 40-90.

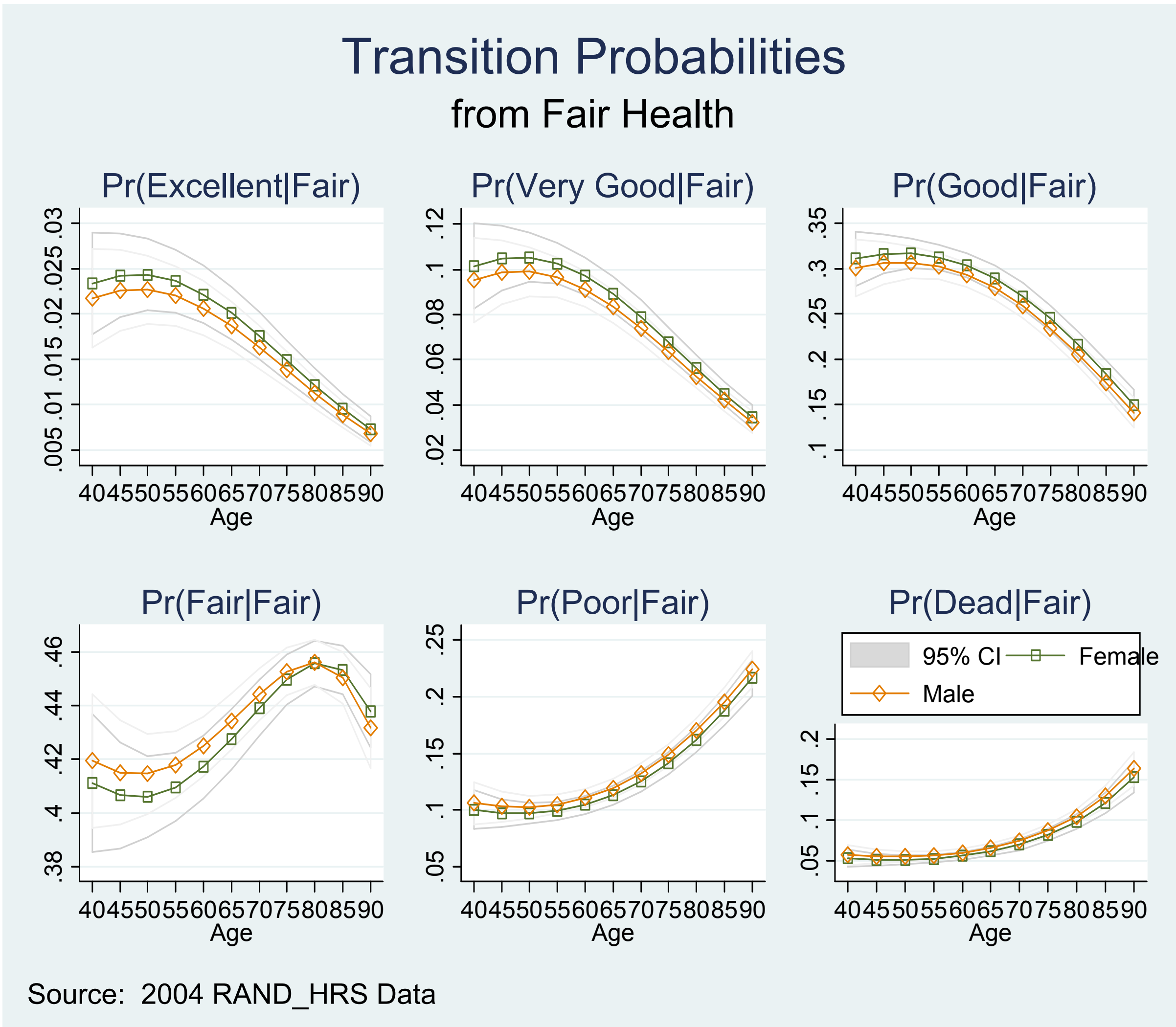

Figure 5: Transition Probabilities from Fair Health State to Other Health States. We report the predictions at the median of household income. Predictions based on ordered Logit estimates for agegroup: 40-90.

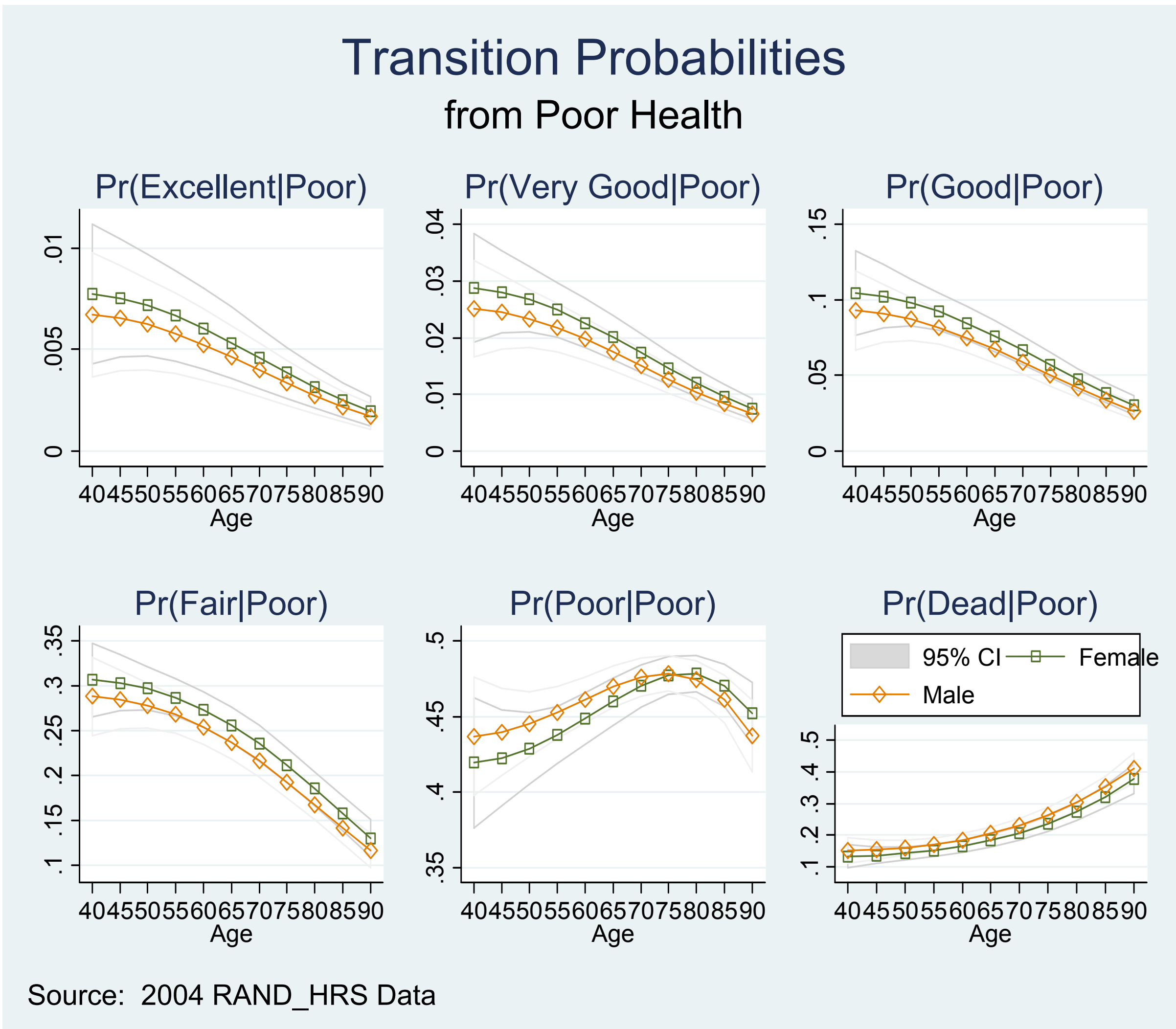

Figure 6: Transition Probabilities from Poor Health State to Other Health States. We report the predictions at the median of household income. Predictions based on ordered Logit estimates for agegroup: 40-90.

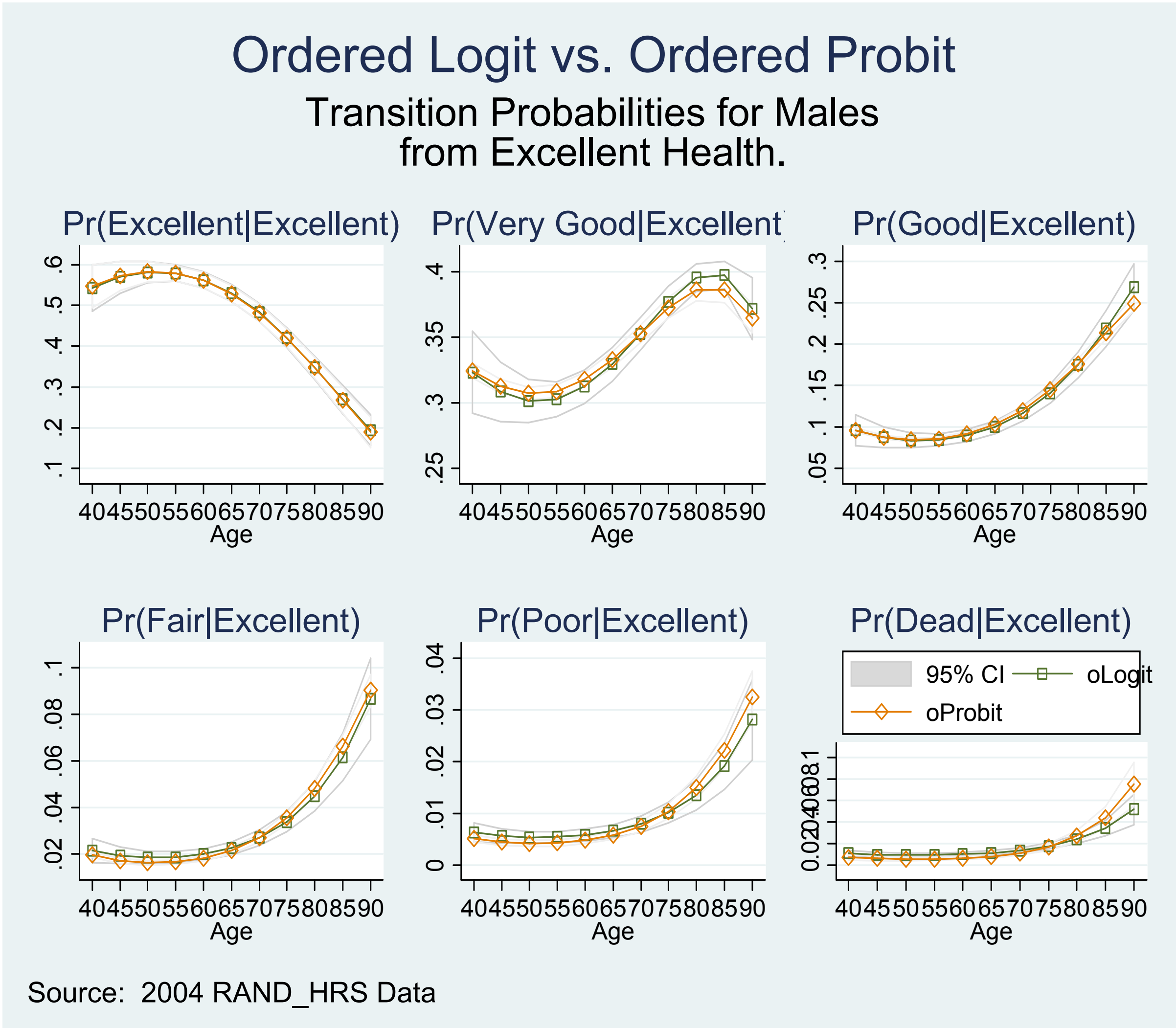

Figure 7: Transition Probabilities from Excellent Health State to Other Health States. We report the predictions at the median of household income. Predictions based on ordered Logit and ordered Probit estimates for agegroup: 40-90.

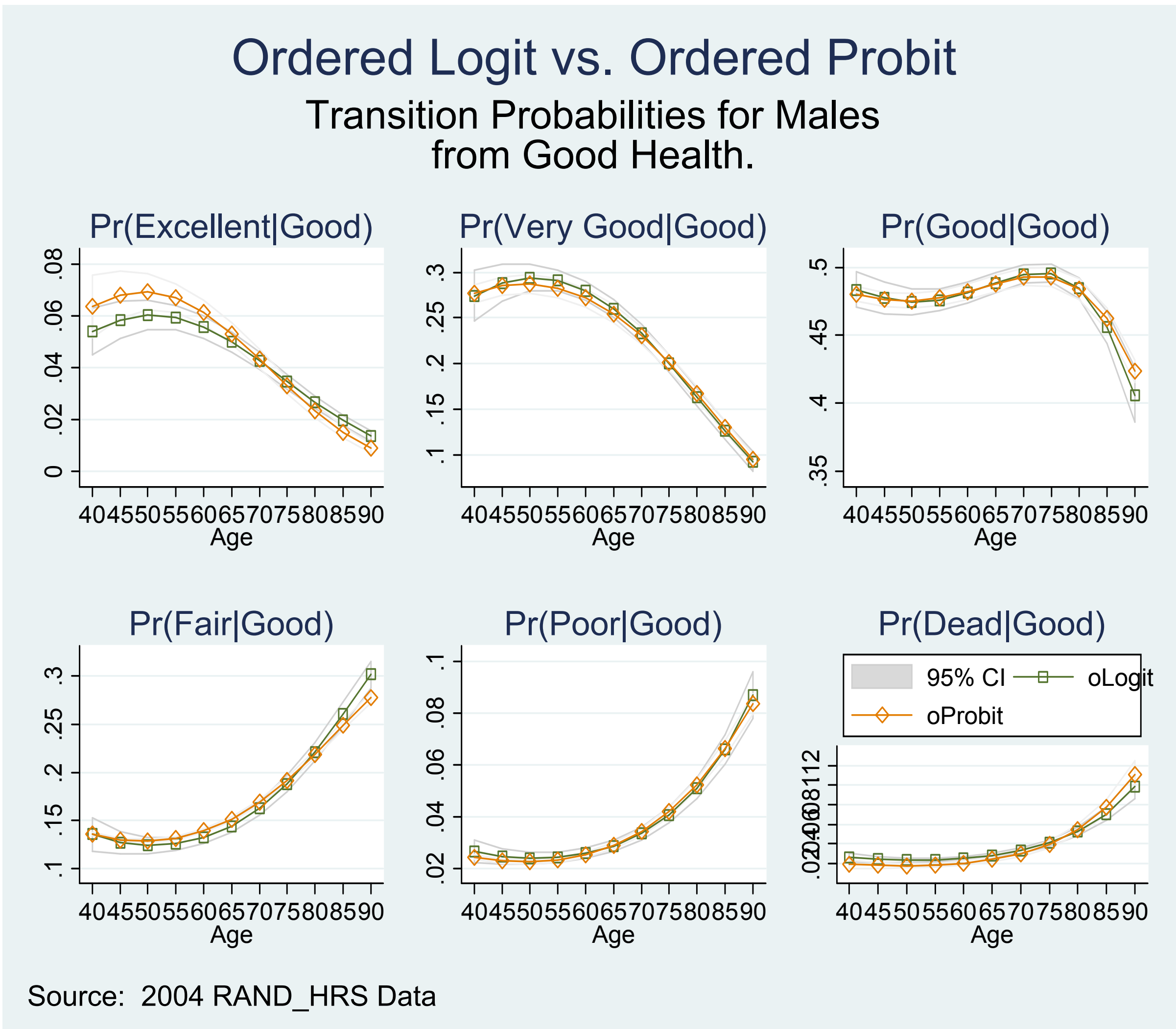

Figure 8: Transition Probabilities from Good Health State to Other Health States. We report the predictions at the median of household income. Predictions based on ordered Logit and ordered Probit estimates for agegroup: 40-90.

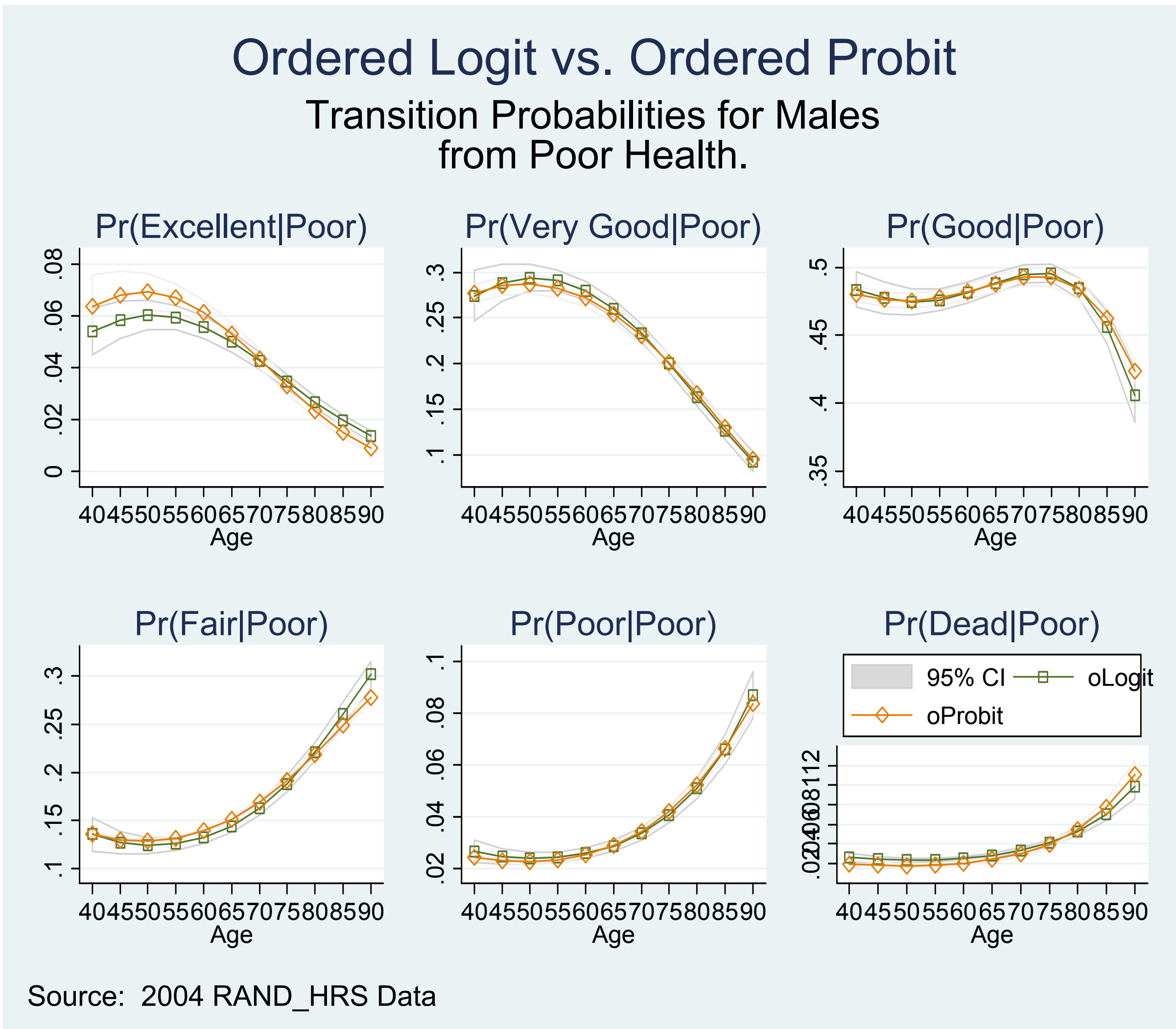

Figure 9: Transition Probabilities from Poor Health State to Other Health States. We report the predictions at the median of household income. Predictions based on ordered Logit and ordered Probit estimates for agegroup: 40-90.

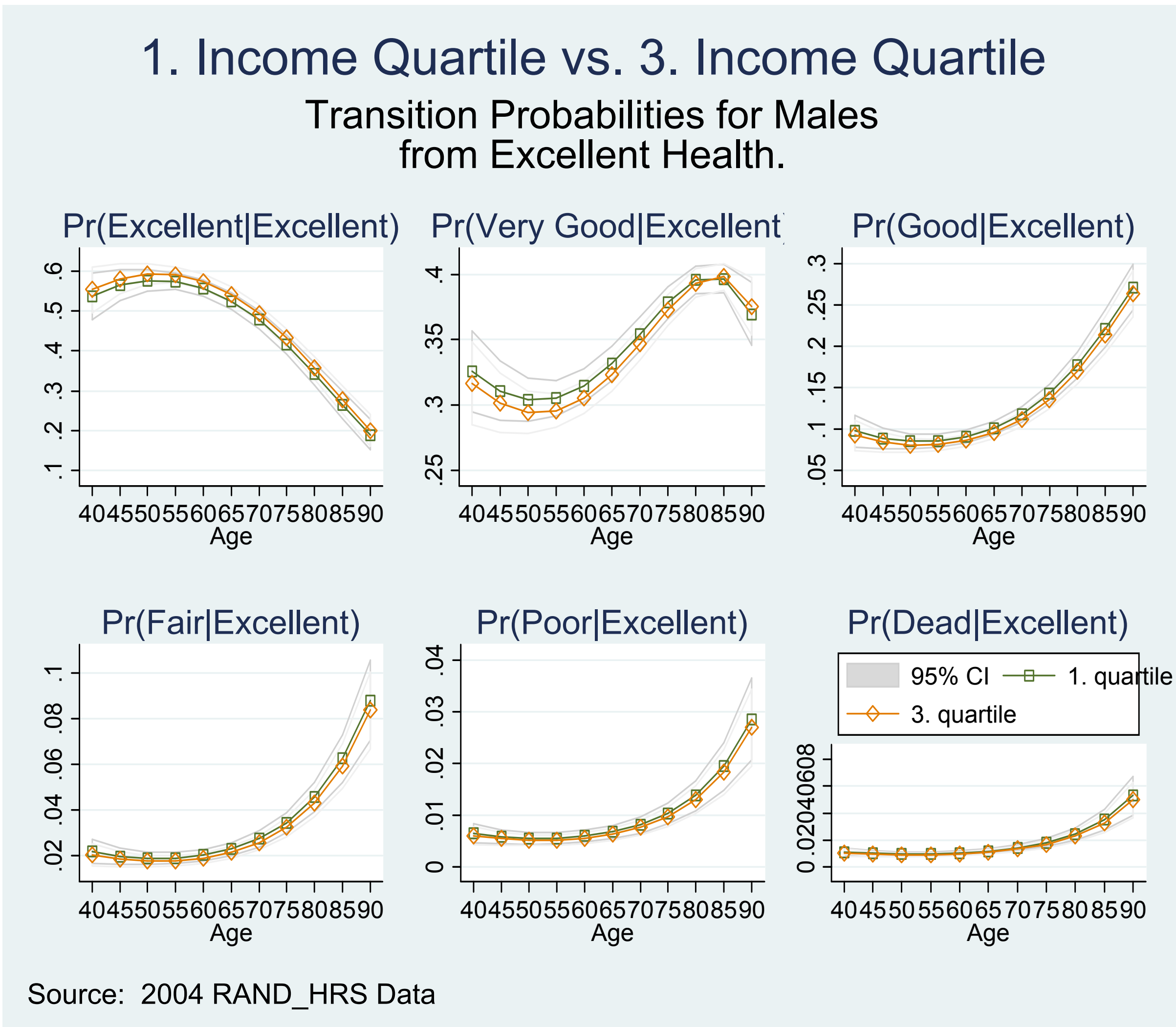

Figure 10: Transition Probabilities for Males from Excellent Health State to Other Health States for 1. and 3. Quartiles. We report the predictions at the median of household income. Predictions based on ordered logit estimates for agegroup: 40-90.

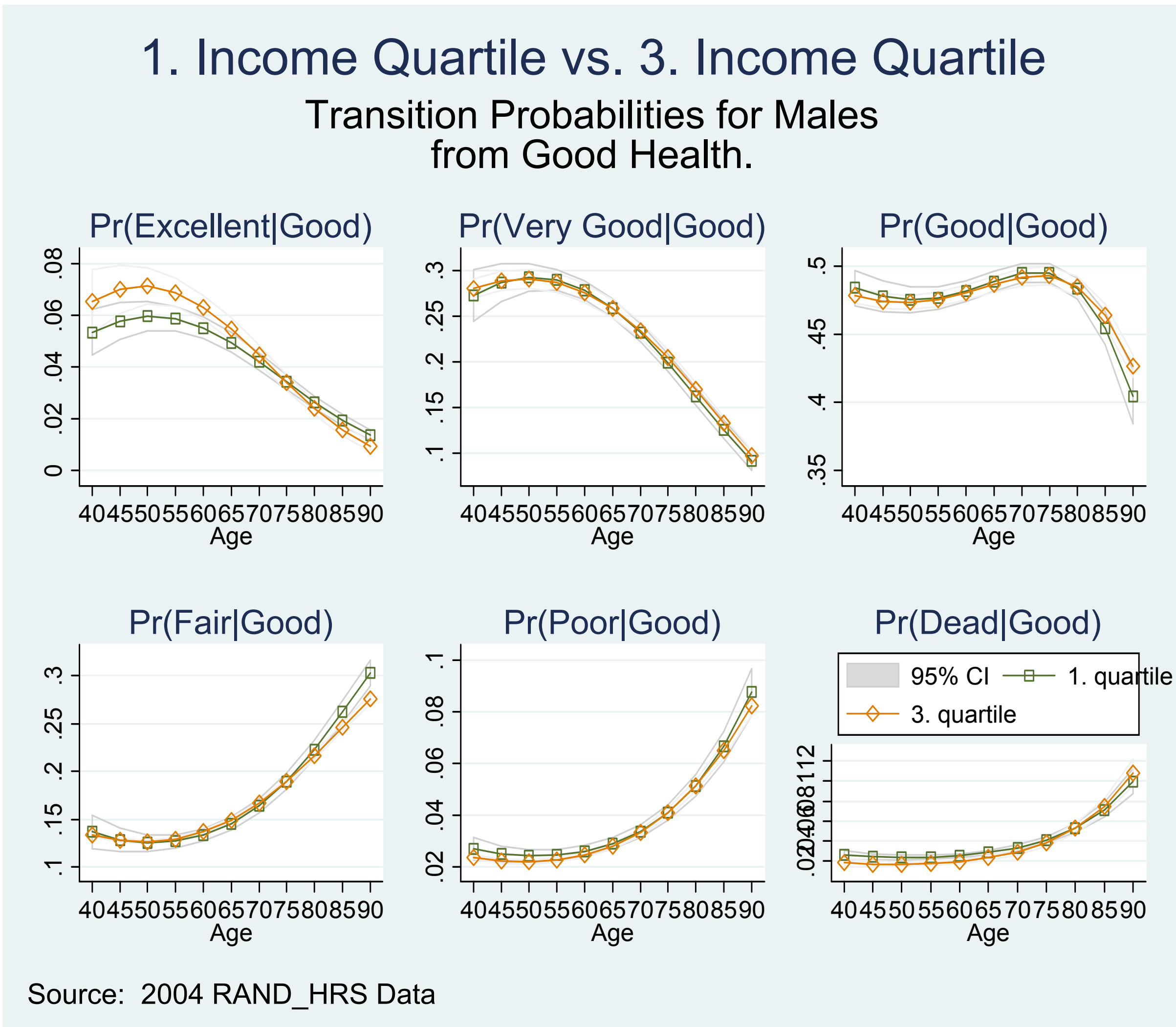

Figure 11: Transition Probabilities for Males from Excellent Health State to Other Health States for 1. and 3. Quartiles. We report the predictions at the median of household income. Predictions based on ordered logit estimates for agegroup: 40-90.

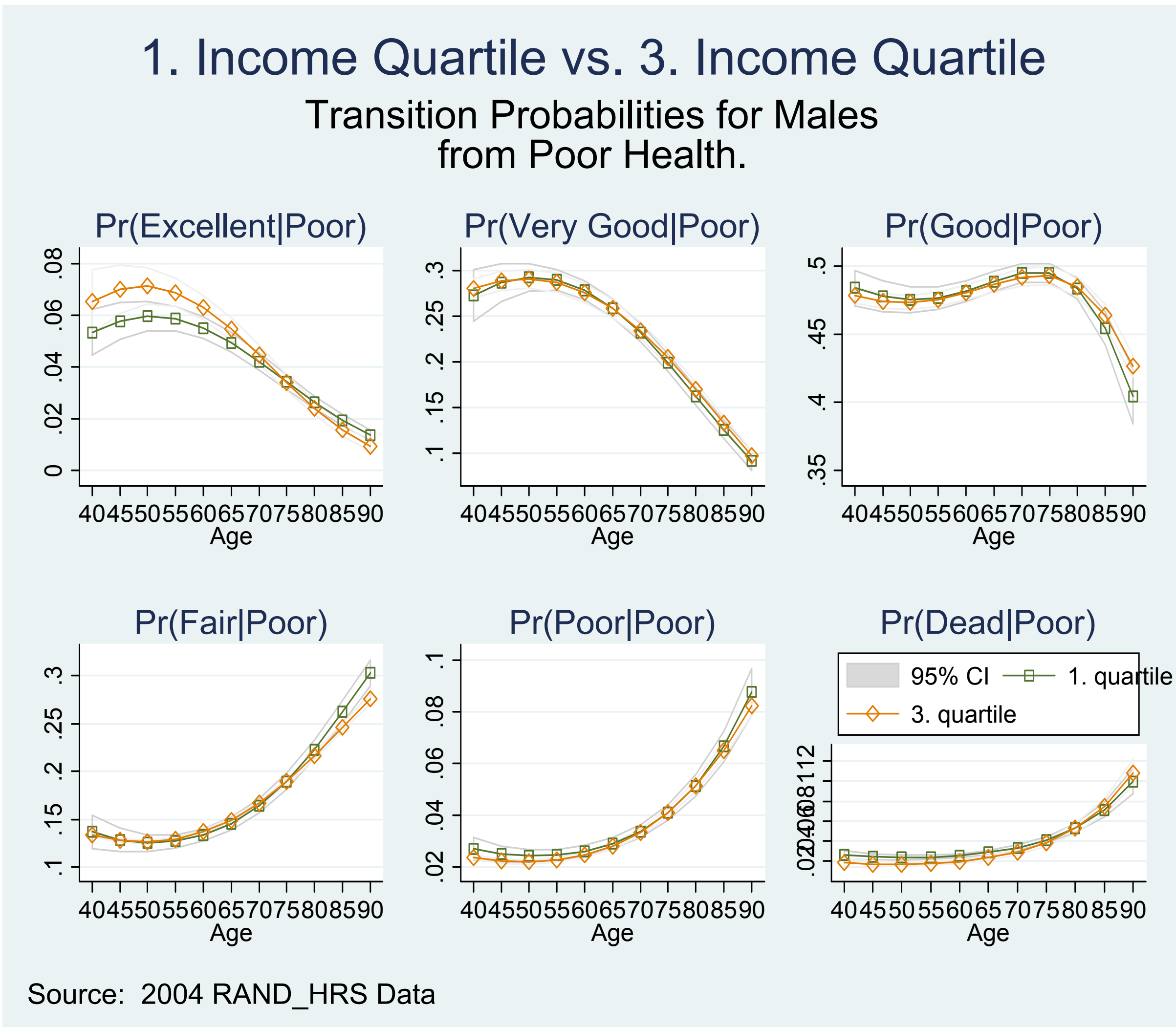

Figure 12: Transition Probabilities for Males from Excellent Health State to Other Health States for 1. and 3. Quartiles. We report the predictions at the median of household income. Predictions based on ordered logit estimates for agegroup: 40-90.

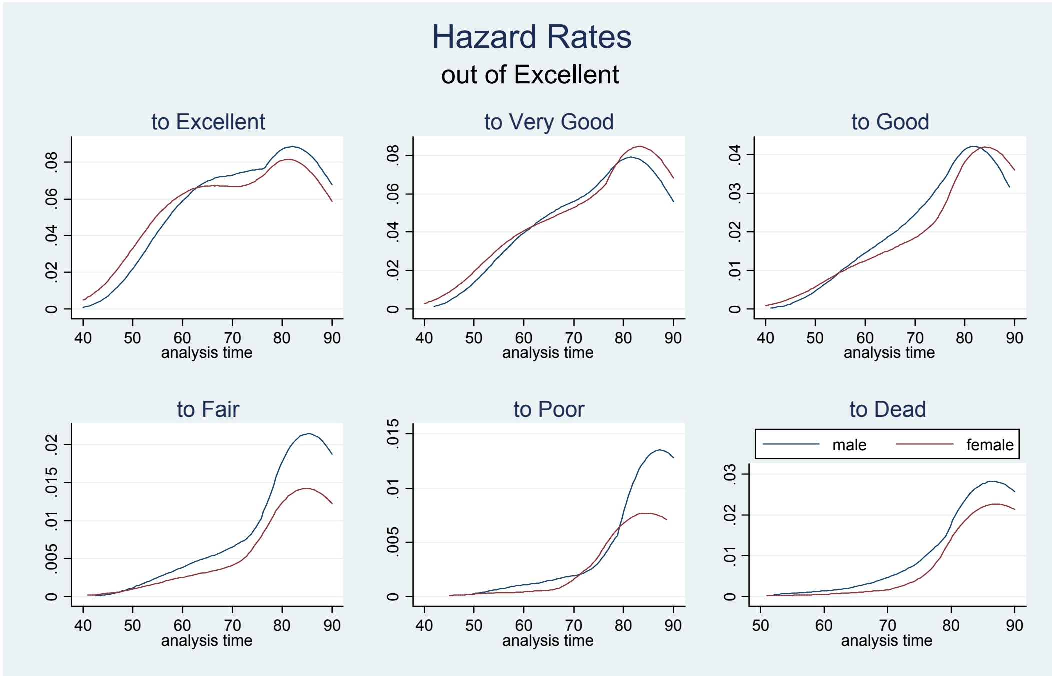

Figure 13: Hazard Rates derived from Non-parametric Nelson-Aalen Estimator from "Excellent" Health State into Various Target Health States. Agegroup: 40-90.

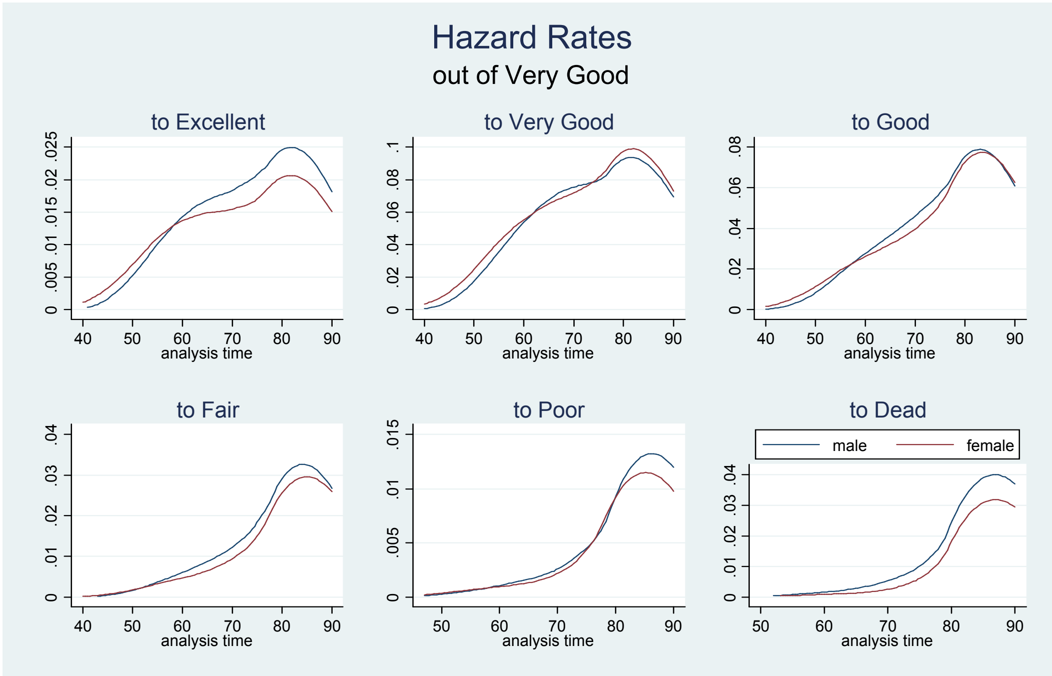

Figure 14: Hazard Rates derived from Non-parametric Nelson-Aalen Estimator from "Very Good" Health State into Various Target Health States. Agegroup: 40-90.

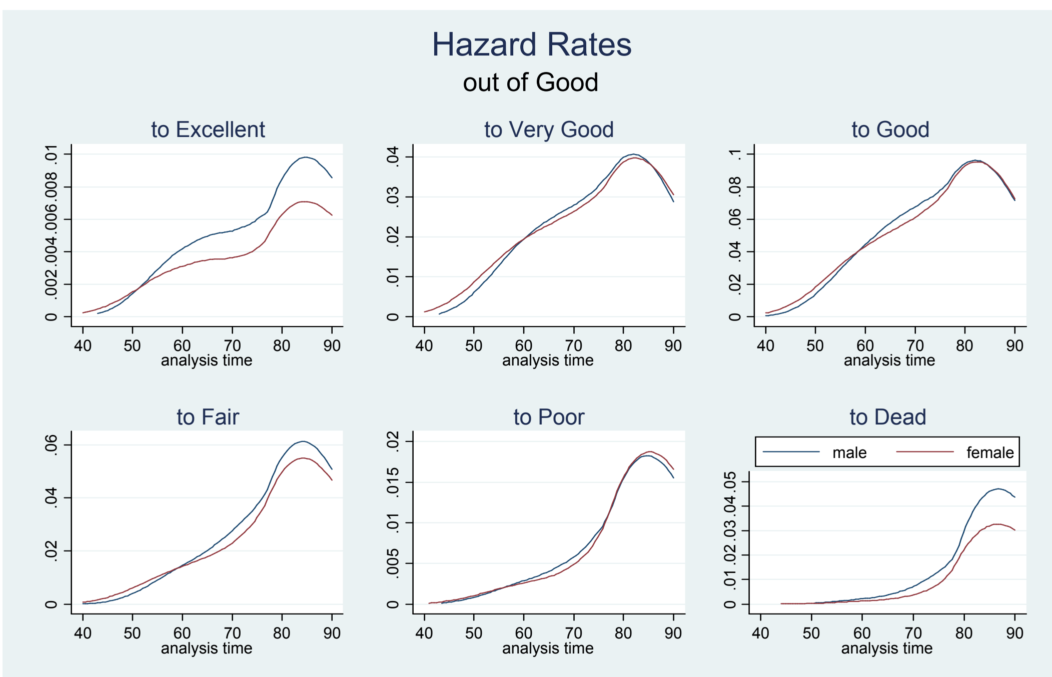

Figure 15: Hazard Rates derived from Non-parametric Nelson-Aalen Estimator from "Good" Health State into Various Target Health States. Agegroup: 40-90.

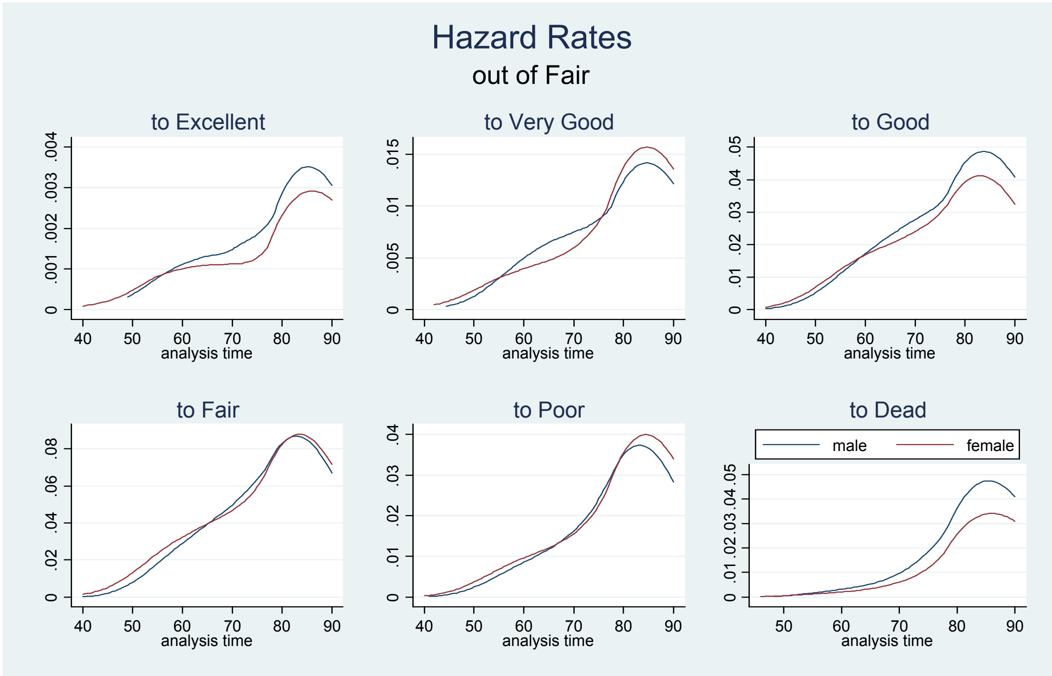

Figure 16: Hazard Rates derived from Non-parametric Nelson-Aalen Estimator from "Fair" Health State into Various Target Health States. Agegroup: 40-90.

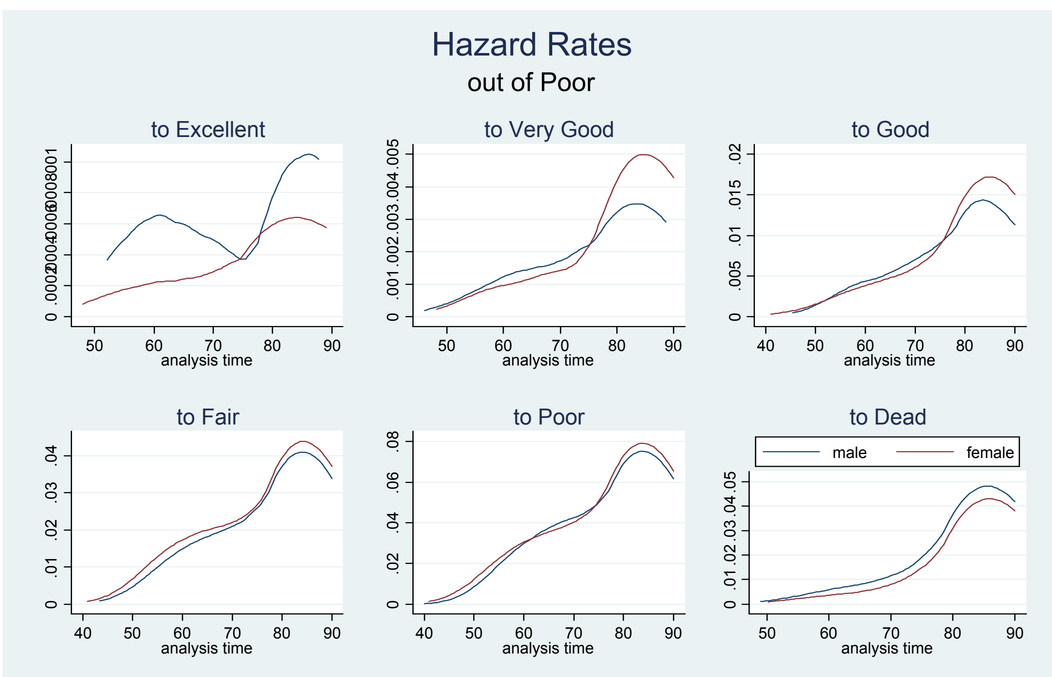

Figure 17: Hazard Rates derived from Non-parametric Nelson-Aalen Estimator from "Poor" Health State into Various Target Health States. Agegroup: 40-90.

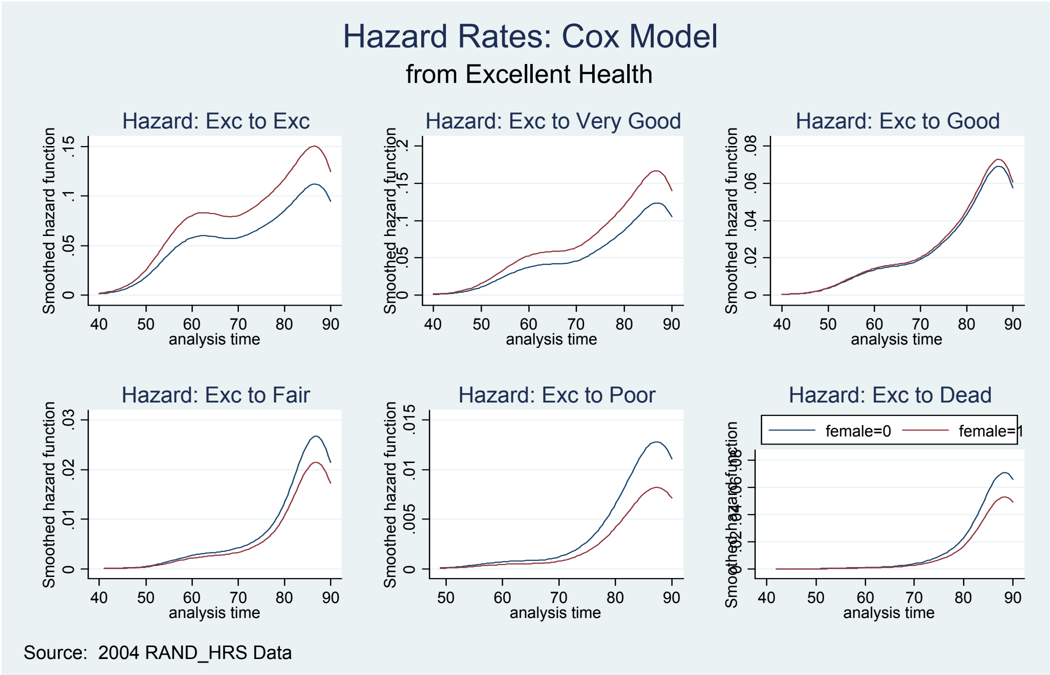

Figure 18: Hazard Rates from Cox Proportional Hazard Model. Initial health state is "Excellent". Agegroup: 40-90.

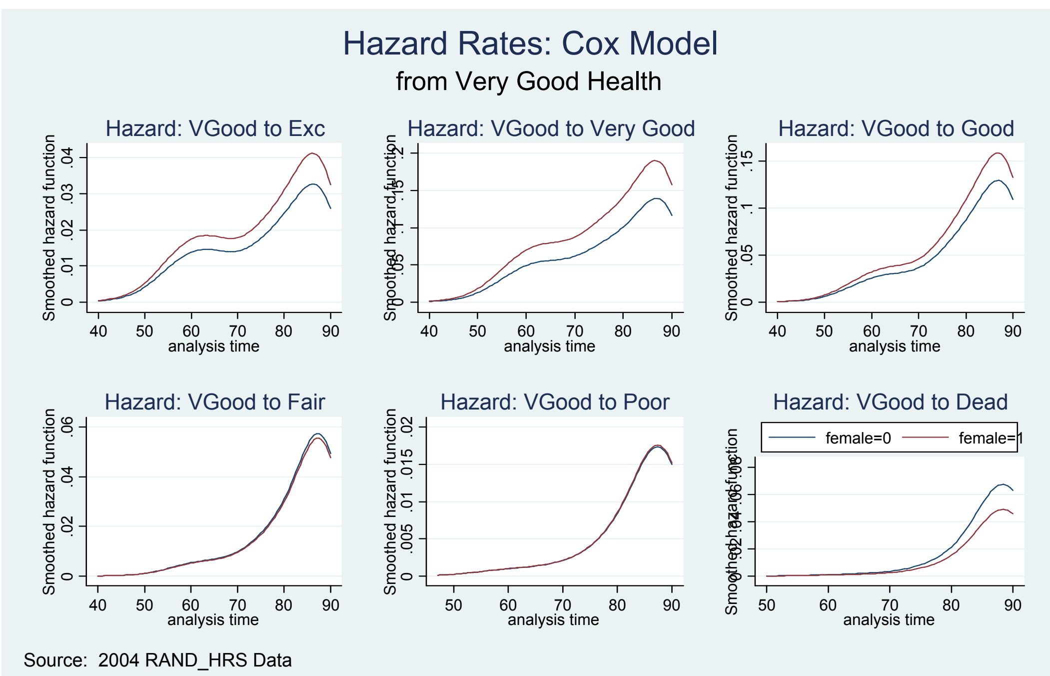

Figure 19: Hazard Rates from Cox Proportional Hazard Model. Initial health state is "Very Good". Agegroup: 40-90.

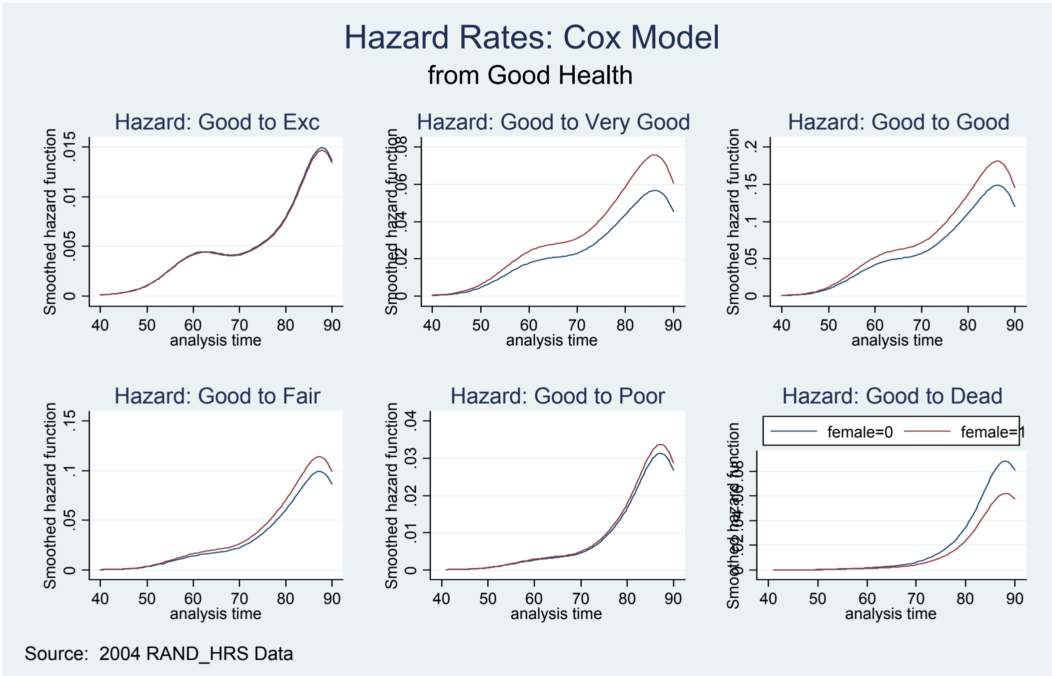

Figure 20: Hazard Rates from Cox Proportional Hazard Model. Initial health state is "Good". Agegroup: 40-90.

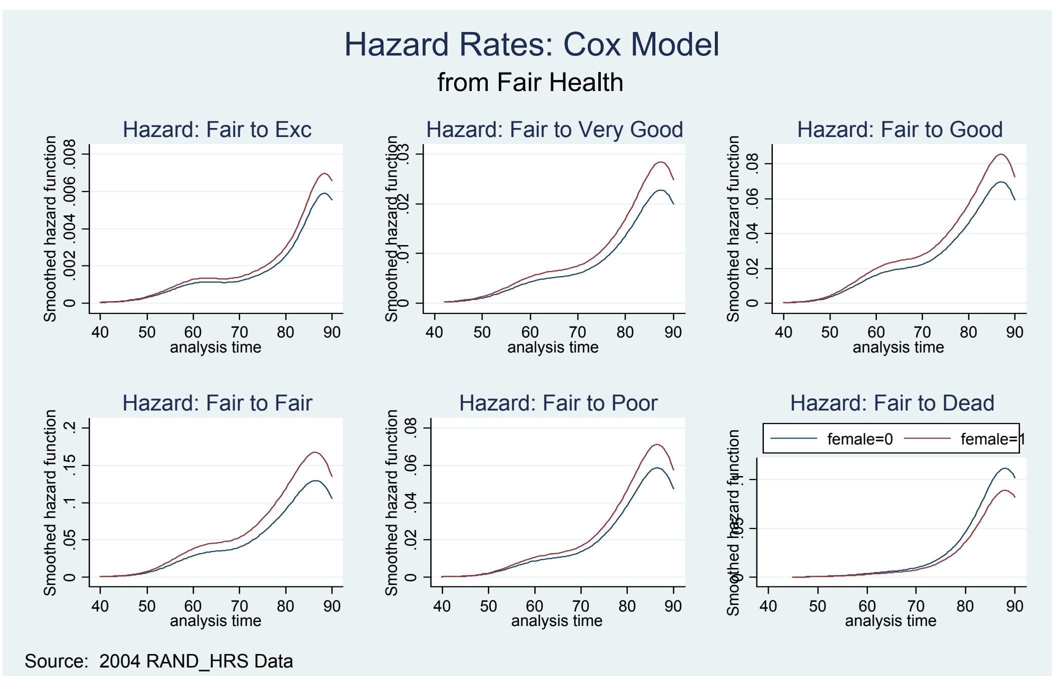

Figure 21: Hazard Rates from Cox Proportional Hazard Model. Initial health state is "Fair". Agegroup: 40-90.

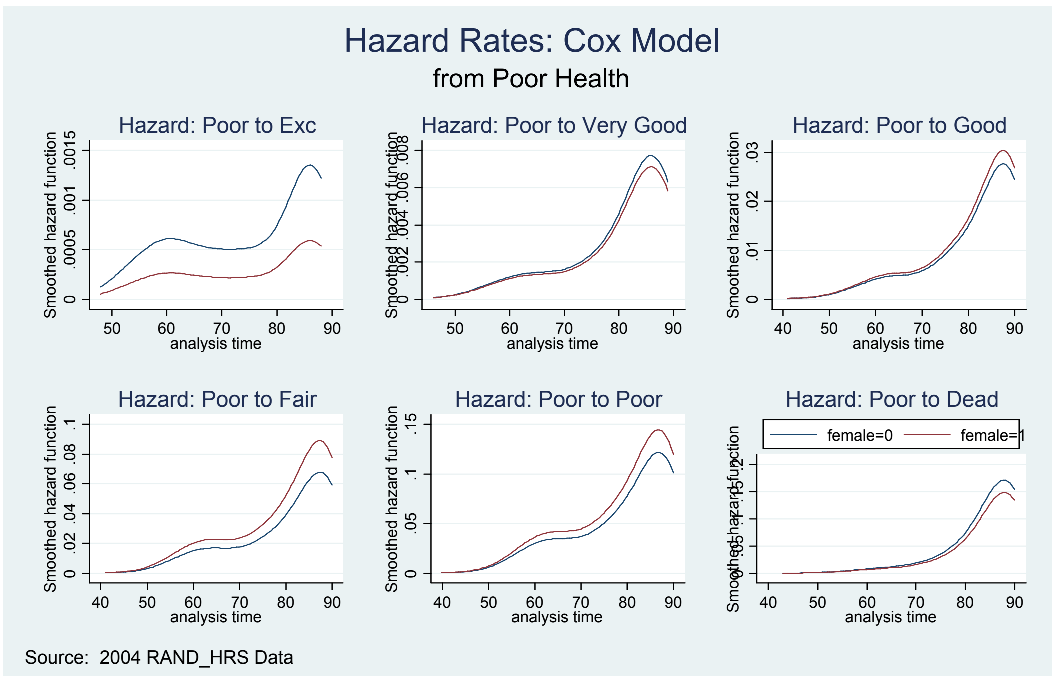

Figure 22: Hazard Rates from Cox Proportional Hazard Model. Initial health state is "Poor". Agegroup: 40-90.