**Appendices 1 to 4, “The Role of Health in Retirement”**

**Alan L. Gustman and Thomas L. Steinmeier**

**February, 2014**

Appendix 1: Construction of the Health and Related Indices

 In this section we examine descriptive statistics for the components of the health index and estimate an IRT (Item Response Theory) model to derive an index from these components. Next, the medical conditions are described and the frequencies of the conditions are noted. The same exercise is done for the personal characteristics which affect retirement. Among the personal characteristics is cognition, for which a set of indicator variables is discussed and a model deriving a cognitive index from these variables is estimated. For all of the variables in the health model, the data set is all the cohorts in the first nine waves of the Health and Retirement Study.

*A. The health index.*

 The first issue to decide what variables to use in constructing the health index. On the one hand are functional limitations, such as the inability to climb a flight of stairs or raise your arms above your head. On the other hand are various medical conditions, such has cancer, diabetes, or heart problems. It would seem that limitations are also likely to play a more important role in retirement decisions than conditions. If a person had difficulty getting around, it would seem to be of secondary importance with respect to the retirement decision whether that difficulty was due to diabetes or to the aftermath of a stroke.

 However, the presence of medical conditions might make a substantial difference to mortality and to the subsequent course of the health variable. Thus, among individuals with a similar current health status measure, those with cancer might be expected to be more likely to die in subsequent surveys than would those who have survived a heart attack, and those with diabetes are more likely to suffer a reduction in the health index than those with hypertension. Hence, our approach is to use the functional limitations as important factors in constructing the health index and to use medical conditions separately as variables to explain mortality and transitions among various levels of the health index.

 The health index we construct utilizes five dimensions of health. They are:

1. Self reported health. This variable has five categories: excellent, very good, good, fair, and poor.

2. Lower body mobility. This variable looks at the answers to seven questions. All seven have the form of “Because of a health problem do you have any difficulty with …?” The seven activities are: (1) walking several blocks, (2) walking one block, (3) sitting for about two hours, (4) getting up from a chair after sitting for long periods, (5) climbing several flights of stairs without resting, (6) climbing one flight of stairs without resting, and (7) stooping, kneeling, or crouching. The values of this variable range from zero to seven and are simply the number of these questions to which the respondent answers “yes.” If the respondent reports that he or she can’t or won’t do the activity, that response is taken to be a “yes.”

3. Upper body agility. This variable is similar to the previous lower body mobility variable and looks at the answers to four questions. The activities these questions inquire about are: (1) reaching or extending your arms above shoulder level, (2) pulling or pushing large objects like a living room chair, (3) lifting or carrying weights over 10 pounds, like a heavy bag of groceries, and (4) picking up a dime from a table. The values of this variable range from zero to four and are simply the number of affirmative responses for these questions.

4. Pain. This variable measures the amount of pain most of the time. It has four values: no pain, mild pain, moderate pain, and severe pain.

5. ADL’s. This variable counts the number of ADLs. The five ADLs considered are: dressing, including putting on shoes and socks, (2) walking across a room, (3) bathing or showering, (4) eating, such as cutting up your food, and (5) using the toilet, including getting up and down. The values of this variable range from zero to five.

*Note that the health index does not include measures of medical conditions, which are included separately.*

 Appendix Table 1.1 gives the distributions of each of these five variables. There were relatively few missing observations on these variables, and the number of observations was typically around 147,000. About 8.7 percent of the observations registered the best category in all five dimensions, while less than 0.09 percent were in the worst category in all five dimensions. The most common problems were problems with mobility, with 63% having difficulty doing at least one mobility activity. Less than half reported agility limitations, pain, or an ADL. Not surprisingly, the least common problems were the ADL’s, which entail somewhat greater limitations than the mobility or agility measures. For the self-reported health measure, the majority of the respondents were in the middle categories as opposed to being either in excellent health or poor health.

 Appendix Table 1.2 gives the correlation matrix among these five variables, assuming that the non-numeric categories are converted to a numeric scale. Perhaps not surprisingly, the most highly correlated variables are mobility and agility. Also not surprisingly, the next most highly correlated variables are these two with the ADL variable. All of these correlations are positive, as expected, but all are substantially less than unity. This means that they are generally measuring somewhat different dimensions of health.

 Unfortunately, in the full health and retirement model, health is a state variable, and the computational burden goes up exponentially as the number of state variables increases. The consequence is that in order to make the model tractable, it is necessary to summarize these measures of health in a single health index, rather than include them all as separate variables in the model.

 There are several approaches one could try to build a health index. For instance, one could try principal components, but the explanatory variables for the analysis are largely categorical in this case, and principal components works best with variables with a quantitative dimension. Running a regression using the health measures on the right hand side runs into the problem that the left hand variable, the underlying health status, is unobservable. We adopt the strategy used by Soldo, Mitchell et al. (2007) in applying item response theory, or IRT for short. IRT can consistently include variables that are neither binary nor have a natural numeric scale, but are categorical (e.g., no pain, mild pain, moderate pain, or severe pain) in a way that principal components cannot. In addition, IRT provides a little more elegant theoretical underpinning than does principal components, which largely lacks a theoretical underpinning for its analysis. The analysis here will use a version of IRT to condense the information in the five health measures into a single variable.

 Item response theory begins with an underlying and unobserved health variable θ, which is presumed to be normally distributed the population. Since this variable has no observable values, we can without loss of generality take its mean to be zero and its standard deviation to be one. To maintain uniformity with the observed health measures, we will take low values for the underlying health variable to correspond to good health and high values to correspond to poor health.

For each of the observed health measures, the probability of falling into one of the response groups is given by an ordered probit, influenced by the value of the underlying health measure. For the observed health measure i, this results in the following probabilities for the response groups:

 Pr(zi = 1) = F(βi,1 - αi θ) for the first response group

 Pr(zi = si) = 1 – F(βi,si-1 - αi θ) for the last response group

 and Pr(zi = j) = F(βi,j - αi θ) – f(βi,j-1 - αi θ) for the other response groups

where the βi,j’s are strictly increasing in their second subscript. The value of αi measures the sensitivity of this health measure to the underlying health variable.

Graphically, we may picture the probabilities for the various response groups as areas under the normal distribution. If the value of θ is zero, or if the value of αi is zero, the situation may be pictured as follows:

βi1

βi2

βi3

βi4

B

C

A

D

E

This figure corresponds to a health measure with five ordered response categories labeled A, B, C, D, and E. The area of each segment is proportional to the probability that a respondent falls into that response category. Note that there is one less value of β than the number of categories of the observed health variable.

 If the respondent has a negative value of the underlying health variable θ (signifying good health), and assuming αi is positive, then the curve for the observed health measure effectively shifts to the left:

βi1

βi2

βi3

βi4

B

C

A

D

E

The effect of this shift is to raise the probability of low (good health) values of the observed health measures and lower the probability of high (poor health) values. If the α’s of the observed health measures are all positive, then a low value of the underlying health measure raises the probabilities of low values of all the observed health measures. The magnitude of the value of α measures this effect. If α is high, then a small reduction in the underlying health measure will increase the probability of the lowest category by a considerable amount, with the opposite result for a small increase in the underlying health variable. As a result, most respondents will find themselves in either the highest or lowest category for that observed health measure.

 The estimated values for this model are the α’s (one for each observed health measure) and the β’s (one for each response category except the last, for each health measure). For the five observed health measures we have been looking at, there are 5 α’s and 23 β’s. The likelihood value for a particular observation is simply the product of the probabilities of the five outcomes for the observed health measures:



where sm is a vector of the five outcomes for the five observed health measures for observation m. The probabilities on the right side of the summation are simply the probabilities of the response groups given above.

To get the unconditional value of Pr(sm), we must integrate over the possible values of θm.



where f() is the standard normal distribution. The model is estimated with maximum likelihood, with the results given Appendix Table 1.3. The parameters are all estimated with great precision, which is to be expected with the large number of observations.

 The next step is to determine the expected value of the underlying health measure, given the observed health measures for a particular observation. That is, we want the expected value of θ, conditional on the observed values of the health measures



The distribution of the expected values of θ are listed in Appendix Table 1.4.[[1]](#footnote-1) This table also lists the average values of the observed health measures.

 Since it is computationally infeasible to treat health as a continuous variable, we divide it into four regions. The first region, below -0.4, show no more than minor self-reported health problems and/or minor pain problems. For self-reported health, a value of 0 is excellent, a value of 1 is very good, and a value of 2 is good. For pain, a value of 0 is no pain, and a value of 1 is minor pain. This region shows very few problems with mobility or agility limitations or ADL’s. The second region, from -0.4 to 0.3, entails two or fewer mobility limitations, occasional agility limitations, but no ADL’s. The two mobility limitations are generally an inability to walk several blocks and an inability to climb several flights of stairs. The third region, from 0.3 to 1.2, entails more mobility and agility limitations but less than one ADL. At the upper end of the range, individuals are unable to do five of the seven mobility tasks and two of the four agility tasks. The last region, above 1.2, entails multiple ADL’s. We may think of these four regions as being in good health, fair health, poor health, and terrible health. About 35 percent of the sample is in good health and another 30 percent is in fair health. 20 percent is in poor health and 9 percent is in terrible health. Recall that the sample is the entire study, including the original older AHEAD cohort.

*B. Medical Conditions.*

 The HRS consistently asks about eight medical conditions. In each case, the questions are of the form "Has a doctor ever told you that you have ...." The eight conditions considered are: hypertension (high blood pressure), diabetes (high blood sugar), cancer, lung disease (chronic bronchitis or emphysema), heart problems (heart attack, coronary heart disease, angina, or congestive heart failure), stroke, psychiatric problems (including emotional or nervous problems), and arthritis (including rheumatism). Given the nature of the question, it is assumed that once an individual has answered positively in one survey, the answer to that question in all subsequent surveys will be positive.

 Appendix Table 1.5 gives the percentages of observations in which the individual has a positive response to the given condition. Arthritis and hypertension are by far the most common conditions, with percentages of between 50 and 60 percent. Heart conditions affect around a quarter of the observations, and diabetes and psychological problems affect almost 20 percent. Cancer and lung problems occur in around 12 percent of the observations, and the least common is stroke, at around 7 percent.

 Appendix Table 1.6 calculates the correlations among the medical conditions. None of the correlations is electrifying large, but they are all positive. Not surprisingly, the correlations between cancer and the other medical conditions are minimal. Also not surprisingly, the mutual correlations among hypertension, heart problems, and stroke are among the strongest correlations, though even these are not very large. These three are also relatively strongly related to diabetes and, to a lesser extent, arthritis. Lung problems are associated with heart problems and arthritis, and psychiatric problems are somewhat related to hypertension, lung problems, and heart problems.

*C. Personal Characteristics.*

 We investigate the effects of five personal characteristics on medical conditions and the health index. These are smoking, alcohol abuse, obesity, depression, and cognition. The smoking, alcohol and obesity variables are fairly straightforward. For depression, there is a standard depression scale, the so-called CES-D (Center for Epidemiological Studies) scale that is available in all waves of the survey.

 For smoking, the HRS asks about past smoking, whether the respondent is currently smoking, and if so, how much. Previous studies (Soldo et. al., 2007) seem to indicate that current smoking is more important in determining current health than is past smoking, so the smoking variable is taken to be a binary variable indicating whether the respondent is smoking currently. This question is asked in all waves of the survey. About 17 percent of the observations are positive for smoking.

 For alcohol consumption, there are a couple of approaches possible. One is the devise a cutoff based on average drinks per day. The first wave asks about how many drinks, on average, the respondent had per day. The remaining surveys ask how many days per week the individual drinks and, for the days when they do drink, how many drinks they have. This can be used to construct the average drinks per day, which should be roughly consistent across waves. The second approach is to use the so-called CAGE questions. This series of four questions asks whether respondents have felt they should cut down on drinking, whether people have ever annoyed them by criticizing their drinking, whether they have felt guilty about their drinking, and whether they have ever had a drink the first thing in the morning. The CAGE questions were introduced in the third wave of the survey, but since they are retrospective questions, this is not too much of a problem. Soldo et. al. found that the CAGE questions were less powerful in predicting current health than were the drinks per day, so we focus on the latter variable, and we will consider respondents who average two or more drinks per day to have a value of one for the alcohol variable. Only 7 percent of the observations register positive for this drinking variable.

 Obesity is measured by the usual BMI calculation. The HRS asked about weight every wave, but it asked each respondent about height only in the initial wave. The respondent is considered obese if the BMI value is over 30. About 26 percent of the observations for obesity are positive.

 For the depression measure, every wave of the HRS asked a set of questions taken from the CES-D (Center for Epidemiological Studies) form. The CES-D is a set of 20 statements of the form like “During the past week, I was bothered by things that usually don’t bother me.” The respondent is given four response categories: rarely or none of the time, some or a little of the time, occasionally or a moderate amount of the time, and most of all of the time. The response categories are given values of 0 to 3, and the response values are added up, giving a range for the CES-D of 0 to 60. The HRS uses a subset of nine of these statements, but one was dropped in wave 2. In wave 1, the four response categories were used, but in subsequent surveys, the question was whether the statement was true for much of the time during the previous week. To maintain consistency across waves, the upper two response categories in wave 1 are combined to form an affirmative answer, as are the lower two response categories. Since there are nine statements rated on a binary scale, the range for the HRS version of the CES-D scale is 0 to 9.

 For the regular CES-D scale, values over 16 are taken as an indication of depression. According to Lenore Radloff in the journal *Applied Psychological Measurement*, about 21 percent of the population scores above this cutoff point. A similar percentage is achieved for the HRS version of the CES-D scale with a cutoff point of greater than 3. Therefore, we consider any respondent with a score of 4 or more on the HRS version of the CES-D scale to have an indication of depression. Roughly 19 percent of the observations in the HRS scored 4 or more on this scale.

 The cognition measure is somewhat more complicated. In waves of the survey starting in 1996, there are four relatively constant measures of cognition. The first is a word recall exercise. Respondents are given a list of 10 words and immediate asked how many they recall. After a few minutes, they are again asked how many of the original words they can recall. Both recall measures are scored on a scale of 0 to 10, indicating the number of words successfully recalled. The final recall measure we use is simply an average of the immediate and delayed recall measures, rounded up if necessary.

 The next measure is the so-called tics (telephone interview of cognitive status) score, which also has a scale of 0 to 10. The items included in this score are whether the respondent knows the date (month, day, year, day of week), the names of the president and vice-president, whether the respondent can count backwards from 20 to 10, and whether the respondent can identify scissors as the usual tool to cut paper and cactus as the prickly plant that grows in the desert. A tics score of less than five prompts the survey instrument to ask the interviewer whether it would be better to have a proxy respondent.

 The third cognitive measure is the so-called serial sevens, where the respondent is asked to subtract seven from 100 and repeat the process four more times. This item is scored on a 0 to 5 basis, indicating how many of the subtractions are correct. It is possible for the respondent to miss one subtraction, say giving 92 as the first answer, and be correct on subsequent subtractions, say by giving 85 as the second answer, even though 85 would not be the correct answer to subtracting seven from 100 twice.

 The fourth measure is an abbreviated version of the Wechsler Adult Intelligence Scale, or wais. This measure asks the respondent to define five progressively more sophisticated words. Each response is graded as not correct, partially correct, or correct. Partially correct answers are given one point and correct scores are given two points. The overall wais score is the sum of the number of points over the five words and thus ranges from 0 to 10.

 The recall tests and the serial sevens were asked in 1996 and later waves for virtually all respondents who were not being interviewed by proxy. The tics score is computed for everyone in 1996, but only for respondents aged 65 and older in later surveys. The wais score was computed for everyone in 1996. In 1998, 2000, 2002, and 2006 it was computed only for those over age 65 or for first time respondents. In 2004 and 2008 it was again computed only for first time respondents.

 In the 1992 and 1994 waves, none of these cognitive measures appeared. In both surveys, however, an immediate recall and delayed recall test involving 20 words was administered. These results were transformed to a 0 to 10 scale such that the distributions in those years approximately matched the distributions in later years. And in the 1992 wave only, there were seven questions asking the respondent how two things were alike. For example, “In what way are an egg and a seed alike?” Each of these questions was graded on a zero to two scale in much the same way the wais questions were later graded. Again, these results were transformed to a 0 to 10 scale such that the distribution of these results approximated the distributions of the wais variable in later years. For all the variables, if an observation was missing for a particular survey, values were taken from adjacent surveys if possible.

 Appendix Table 1.7 lists the distributions of these variables over all waves of the survey.

Looking at the distributions, the median number of words recalled is six out of ten for the immediate recall and five out of ten for the delayed recall, for a median average of five. The upper tail of the distribution falls off smoothly, suggesting that people writing down the words and later reciting them is not a serious problem. The tics scores are heavily concentrated at the upper end, which is to be expected since this is lowest level cognitive functioning variable. Also, individuals who would otherwise have scored low on the tics were probably being interviewed by proxy. On the 2000 wave and later, the interviewer was asked if the reason for a proxy was that the respondent was cognitively impaired, and we can probably infer a low cognitive score if the answer was affirmative. Prior to 2000, however, there does not appear to be any information as to whether a proxy was used because the respondent was unable to give the interview or was unwilling to do so. In the tics questions, those most frequently missed are day of week (i.e., Tuesday vs. Wednesday) and/or the name of the Vice President.

 The serial sevens show a bimodal distribution. Nearly half of the respondents were able to complete the task correctly, and another fifth missed a single subtraction. The least common result is that the respondent missed exactly three out of the five subtractions; cases where the respondent either did not do any subtractions correctly or did more than two subtractions correctly are more common. The wais scores, in contrast, shows a unimodal distribution, with the modal score being six out of ten.

Appendix Table 1.8 gives the correlations for the four components of the cognition measure. The correlations are generally in the 0.3 to 0.4 range. In general, the correlations involving the tics score are more toward the upper end of this range, and the other correlations, with the exception of the correlation of the serial sevens with the wais score, are toward the lower end of the range. The higher correlations involving the tics score reflect a basic feature of the tics score, namely, if the tics score is low, almost all the other scores are low as well.

The cognition measure is estimated with an item response theory (IRT) model based on the four observed cognition measures. This is very similar in nature to the previous IRT model for the underlying health status. The model is estimated with maximum likelihood, and the estimated coefficients are presented in Appendix Table 1.9. All of the estimated coefficients are highly significant, which is not surprising given the sample size.

The coefficients of the IRT model are not very informative, however. Appendix Table 1.10 gives the distribution of the values of the underlying cognitive status variable, along with the average values of the component cognition measures in 0.1 interval ranges. Higher values of the cognition index reflect higher values of the underlying component measures. For all four observed cognitive measures, the average values increase as the cognition index increases, as they should. For low values of the cognition index, the word recall score, the serial sevens score and the wais score are all fairly low, while the tics score is steadily and rapidly increasing. This reflects the fact that the tics score is a more basic measure of cognitive functioning than the other three measures. At the high end of the cognition index, respondents performed fairly well on the tics measure and the serial sevens measure, while the word recall scores and the wais scores were increasing more rapidly.

For the full health and retirement model, we group individuals according to their cognitive scores. Cognitive indices below -1.2 correspond to tics scores less than 8. Given the nature of the tics measure, tics scores less than 8 suggest at least a mild cognitive impairment. Below a cognitive index of -1.2, the scores on the word recall, serial sevens, and wais measures are also fairly low. At the other end of the range, cognitive indices above 0.5 are associated with word recall scores and wais scores both above 6, and these individuals generally perform very well on the serial sevens measure. As would be expected, individuals in this range have very few problems with the tics measure.

 Using these breaks, it is possible to make the following cognitive groups:

 Cognitive Index

 less than -1.2 Cognitively impaired

 -1.2 to 0.5 Average cognitive abilities

 above 0.5 Excellent cognitive abilities

On this scale, around 8 percent of the observations have some cognitive impairment and 29 percent have excellent cognitive abilities, with the remaining 73 percent having average cognitive abilities.

 For the structural health and retirement model, the personal characteristics are considered to be exogenous, since we don’t really have any theoretical basis for determining them. They also are non-stochastic, since making them stochastic would exacerbate the state space problem. For most individuals and for most of the personal characteristics, the measured characteristics are generally unchanging over time. That is, someone who registered as a smoker in one wave of the survey generally registers as a smoker on most if not all of the other waves as well, and someone who was obese in one wave is generally classified as obese in most if not all of the other waves. For individuals who occasionally changed classifications in the personal characteristic, we assign the value of the personal characteristic to be the value in the majority of the waves where the individual was observed. For the cognitive measure, the individual is considered to be cognitively impaired if he registered as cognitively impaired in a majority of the waves where he was observed, and likewise for cognitive excellence. Otherwise he is considered to be cognitively average.

Appendix Table 1.1

Distribution of the Components of the Health Index

 Excellent 13.3%

 Very Good 28.0

Self-Reported Health Good 30.4

 Fair 19.1

 Poor 9.2

 0 37.1%

 1 16.7

 2 11.8

Number of Mobility 3 9.3

 Limitations 4 7.5

 5 6.6

 6 6.8

 7 4.1

 0 62.2%

Number of Agility 1 15.4

 Limitations 2 12.1

 3 7.7

 4 2.6

 None 70.4%

Level of Pain Mild 8.3

 Moderate 15.9

 Severe 5.4

 0 85.0%

 1 7.1

Number of ADL’s 2 3.1

 3 1.9

 4 1.4

 5 1.5

Appendix Table 1.2

Correlations of the Components of the Health Index

 Mobility Agility Pain ADL’s

SR Health 0.56 0.49 0.41 0.38

Mobility 0.69 0.50 0.57

Agility 0.43 0.58

Pain 0.30

Appendix Table 1.3

Estimates of the Item Response Health Index Model

 Self-Reported Health

* 0.915

 β1 -1.495

 β2 -0.316

 β3 0.768

 β4 1.831

 Mobility Limitations

* 2.033

 β1 -0.776

 β2 0.163

 β3 0.872

 β4 1.522

 β5 2.154

 β6 2.878

 β7 3.998

 Agility Limitations

* 1.561

 β1 0.550

 β2 1.411

 β3 2.387

 β4 3.620

 Pain

* 0.855

 β1 0.684

 β2 1.049

 β3 2.119

 ADL’s

* 1.977

 β1 2.332

 β2 3.195

 β3 3.766

 β4 4.261

 β5 4.833

 All coefficients are have t-values of at least 20.

Appendix Table 1.4

Ranges of the Health Index

 Self

 Health Reported Cumulative

Index Range Health Mobility Agility Pain ADL’s Percent

-1.5 to -1.4 0.00 0.00 0.00 0.00 0.00 8.7

-1.0 to -0.9 1.00 0.00 0.00 0.00 0.00 20.9

-0.9 to -0.8 0.00 0.00 0.00 1.00 0.00 21.1

-0.8 to -0.7 1.97 0.00 0.00 0.03 0.00 29.2

-0.7 to -0.6 1.00 0.00 0.00 1.00 0.00 29.8

-0.6 to -0.5 1.40 0.39 0.08 0.22 0.00 34.1

-0.5 to -0.4 2.00 0.00 0.00 1.00 0.00 34.7

---------------------------------------------------------------------------------------------------------

-0.4 to -0.3 1.16 0.72 0.16 0.22 0.00 40.3

-0.3 to -0.2 1.85 0.88 0.13 0.19 0.00 45.7

-0.2 to -0.1 1.48 1.39 0.16 0.29 0.00 49.8

-0.1 to 0.0 1.78 1.38 0.36 0.43 0.00 54.2

 0.0 to 0.1 1.97 1.72 0.36 0.48 0.00 58.2

 0.1 to 0.2 1.68 2.24 0.50 0.48 0.00 61.8

 0.2 to 0.3 2.12 2.10 0.75 0.64 0.01 65.2

 ---------------------------------------------------------------------------------------------------------

 0.3 to 0.4 2.14 2.62 0.82 0.61 0.03 68.5

 0.4 to 0.5 2.26 2.93 0.86 0.88 0.04 71.6

 0.5 to 0.6 2.27 3.20 1.05 0.97 0.11 74.6

 0.6 to 0.7 2.33 3.53 1.29 0.92 0.17 77.6

 0.7 to 0.8 2.57 3.88 1.41 1.04 0.23 80.4

 0.8 to 0.9 2.75 4.15 1.58 1.11 0.33 82.6

 0.9 to 1.0 2.64 4.46 1.74 1.30 0.41 84.3

 1.0 to 1.1 2.71 4.72 1.91 1.26 0.54 86.2

 1.1 to 1.2 2.86 5.03 2.05 1.32 0.70 88.4

---------------------------------------------------------------------------------------------------------

 1.2 to 1.3 2.93 5.23 2.21 1.46 1.06 90.6

 1.3 to 1.4 3.07 5.57 2.34 1.56 1.32 92.3

 1.4 to 1.5 3.09 5.89 2.53 1.54 1.50 93.6

 1.5 to 1.6 3.18 6.01 2.54 1.64 1.97 94.7

 1.6 to 1.7 3.31 6.10 2.68 1.62 2.30 95.7

 1.7 to 1.8 3.33 6.31 2.84 1.70 2.54 96.7

 1.8 to 1.9 3.44 6.42 3.07 1.89 2.83 97.7

 1.9 to 2.0 3.47 6.50 3.26 1.76 3.53 98.5

 2.0 to 2.1 3.63 6.65 3.37 1.81 3.70 99.0

 2.1 to 2.2 3.70 6.81 3.40 1.97 3.94 99.3

 2.2 to 2.3 3.84 6.77 3.56 2.22 4.08 99.5

 2.3 to 2.4 3.80 6.84 3.78 2.54 4.14 99.7

 2.4 to 2.5 3.86 7.00 3.81 1.73 4.70 100

 2.5 to 2.6 3.42 7.00 4.00 2.16 5.00 100

Appendix Table 1.5

Frequency of Medical Conditions

 Percent

 with Condition

 Hyperten 53.3%

 Diabetes 17.8

 Cancer 12.4

 Lungs 11.7

 Heart 23.8

 Stroke 6.9

 Psych 18.7

 Arthritis 58.2

Appendix Table 1.6

Correlations Among Medical Conditions

 Diabetes Cancer Lungs Heart Stroke Psych Arthritis

Hypertension 0.217 0.051 0.071 0.196 0.142 0.110 0.185

Diabetes 0.036 0.057 0.151 0.109 0.083 0.112

Cancer 0.064 0.080 0.039 0.038 0.087

Lungs 0.163 0.071 0.147 0.129

Heart 0.185 0.128 0.169

Stroke 0.087 0.081

Psych 0.155

Appendix Table 1.7

Scores on Components of Cognition Measure

 Score

 0 1 2 3 4 5 6 7 8 9 10

Recall 0.5% 2.0 5.5 10.1 16.3 21.8 20.8 13.2 6.4 2.5 0.8

Tics 0.1 0.1 2.9 0.3 0.5 0.9 1.6 3.6 9.5 25.3 55.2

Ser7s 11.0 9.3 7.7 11.8 18.2 42.0

Wais 2.2 2.3 4.0 8.9 13.1 18.6 19.1 17.0 9.4 4.3 1.2

Appendix Table 1.8

Correlations of the Components of the Cognition Measure

 tics ser7s wais

 recall 0.389 0.333 0.308

 tics 0.427 0.345

 ser7s 0.374

Appendix Table 1.9

Estimates of the Item Response Cognition Index Model

 Word Recall tics

 α 0.693 α 0.958

 β1 -3.148 β1 -4.416

 β2 -2.39 β2 -3.948

 β3 -1.689 β3 -3.562

 β4 -1.072 β4 -3.188

 β5 -0.445 β5 -2.816

 β6 0.232 β6 -2.398

 β7 0.928 β7 -2.178

 β8 1.576 β8 -1.812

 β9 2.195 β9 -1.186

 β10 2.838 β10 -0.168

 Serial Sevens Wais

 α 0.956 α 0.738

 β1 -1.649 β1 -2.521

 β2 -1.081 β2 -2.097

 β3 -0.731 β3 -1.675

 β4 -0.283 β4 -1.116

 β5 0.334 β5 -0.582

 β6 0.017

 β7 0.619

 β8 1.307

 β9 1.975

 β10 2.756

 All coefficients have a t-value of at least 35 with the exception

 of β6 for the wais score, which has a t-value of 4.1.

Appendix Table 1.10

Ranges of the Cognition Index

 Cognition Cumulative

Index Range recall tics ser7s wais Percent

-3.4 to -3.3 0.00 0.00 0.00 0.00 0.0

-3.2 to -3.1 0.50 0.36 0.00 0.14 0.0

-3.1 to -3.0 0.00 0.00 0.00 2.00 0.0

-3.0 to -2.9 0.30 1.70 0.00 0.00 0.0

-2.9 to -2.8 0.81 1.21 0.05 0.14 0.1

-2.8 to -2.7 0.73 1.98 0.00 0.50 0.1

-2.7 to -2.6 0.94 2.65 0.02 0.43 0.2

-2.6 to -2.5 0.71 3.21 0.01 1.05 0.3

-2.5 to -2.4 1.30 3.23 0.01 0.95 0.4

-2.4 to -2.3 1.51 3.75 0.00 1.08 0.5

-2.3 to -2.2 1.54 4.44 0.03 1.14 0.7

-2.2 to -2.1 1.80 4.90 0.05 1.24 1.0

-2.1 to -2.0 1.90 5.39 0.05 1.48 1.3

-2.0 to -1.9 2.18 5.96 0.08 1.53 1.7

-1.9 to -1.8 2.42 6.12 0.11 1.87 2.2

-1.8 to -1.7 2.33 6.47 0.19 2.27 2.8

-1.7 to -1.6 2.59 6.84 0.28 2.36 3.5

-1.6 to -1.5 2.89 6.93 0.45 2.39 4.2

-1.5 to -1.4 3.04 7.49 0.48 2.43 5.2

-1.4 to -1.3 3.13 7.57 0.60 3.00 6.4

-1.3 to -1.2 3.32 7.76 0.78 3.23 7.8

------------------------------------------------------------------------------------------------------------------

-1.2 to -1.1 3.33 8.06 0.92 3.56 9.3

-1.1 to -1.0 3.49 8.27 1.19 3.55 11.1

-1.0 to -0.9 3.61 8.49 1.41 3.54 12.9

-0.9 to -0.8 3.78 8.53 1.60 3.83 14.8

-0.8 to -0.7 3.98 8.66 1.82 3.95 17.2

-0.7 to -0.6 4.10 8.67 2.25 4.18 20.1

-0.6 to -0.5 4.30 8.83 2.41 4.41 23.3

-0.5 to -0.4 4.36 9.03 2.58 4.60 27.2

-0.4 to -0.3 4.53 9.16 2.84 4.80 31.5

-0.3 to -0.2 4.62 9.38 3.00 4.97 36.2

-0.2 to -0.1 4.95 9.47 3.15 5.04 40.7

-0.1 to 0.0 5.00 9.27 3.72 5.19 44.2

 0.0 to 0.1 5.00 9.44 3.87 5.29 49.3

 0.1 to 0.2 5.24 9.61 3.86 5.65 55.8

 0.2 to 0.3 5.75 9.70 3.79 5.95 60.8

 0.3 to 0.4 5.40 9.79 4.36 5.71 65.9

 0.4 to 0.5 5.45 9.73 4.69 5.84 71.0

Appendix Table 1.10 (cont.)

 Cognition Cumulative

Index Range recall tics ser7s wais Percent

------------------------------------------------------------------------------------------------------------------

 0.5 to 0.6 6.68 9.68 4.04 6.81 74.3

 0.6 to 0.7 5.86 9.94 4.66 6.20 80.1

 0.7 to 0.8 5.97 9.92 4.92 6.44 85.2

 0.8 to 0.9 7.14 9.74 4.21 8.07 86.2

 0.9 to 1.0 6.46 9.97 4.90 6.93 91.1

 1.0 to 1.1 8.11 9.97 4.82 6.38 91.9

 1.1 to 1.2 6.46 9.98 4.95 7.72 94.8

 1.2 to 1.3 7.71 10.00 4.96 7.40 96.7

 1.3 to 1.4 5.97 9.99 4.96 9.18 97.4

 1.4 to 1.5 7.94 10.00 4.99 8.11 98.7

 1.5 to 1.6 6.00 10.00 5.00 10.00 98.9

 1.6 to 1.7 8.62 10.00 4.99 8.40 99.4

 1.7 to 1.8 7.00 10.00 5.00 10.00 99.6

 1.8 to 1.9 9.35 10.00 5.00 8.65 99.8

 1.9 to 2.0 8.00 10.00 5.00 10.00 99.9

 2.0 to 2.1 10.00 10.00 5.00 9.00 99.9

 2.1 to 2.2 9.00 10.00 5.00 10.00 100

 2.4 to 2.5 10.00 10.00 5.00 10.00 100

Appendix 2: Detailed Estimates of the Health Model.

 In this section, transition equations for the various health states are estimated. These include estimates of mortality, transitions into the various medical conditions, and transitions among the four health categories. Adjustments necessary to produce dynamically sensible results are determined, and summary statistics for the overall health sub-model are presented.

*A. Mortality.*

 Mortality is taken to depend on health status, medical conditions, personal characteristics, and age. Because of this dependence, we cannot simply use published mortality tables from the census or other publicly available sources, but rather mortality must be estimated from the HRS. There are issues of deriving mortality from the HRS, however. When an individual who was interviewed in one survey is not able to be interviewed in the next survey, an effort is made to ascertain why. Some individuals simply refuse but are known to be alive, while in other cases the interviewers are told that the individual has died. For individuals who simply disappear, the HRS uses the National Death Index to ascertain whether the cause of non-interview is death. However, it is known that this procedure probably does not capture all the non-interviewees who have died.

 Given that it is likely that the HRS probably underestimates actual mortality, the approach used in this project is to estimate the relationship between mortality and health status, medical conditions, and personal characteristics from the HRS. When this relationship is used in the dynamic simulations, the overall level of mortality by age is then adjusted to match the aggregate mortality rate as reported from other sources. This procedure should preserve the differences in relative mortality among the various groups while giving overall results that match the actual mortality experience.

 The mortality hazard is given by the functional form:

 m = 0.01 e βX

where m is the mortality hazard and X is a list of variables influencing mortality. The X variables include a constant, age – 65, (age – 65)2, and interactions between these three terms and each of the health status, medical conditions, and personal characteristics variables.

 The results of estimating this mortality relationship are presented in Appendix Table 2.1. Most of the health status, medical conditions, and personal characteristics have at least some significant effect on mortality. About half of the medical conditions have effects that vary with age and/or age2. The signs of the coefficients are generally as expected. The more serious medical conditions and health problems, as well as smoking, all seem to have an elevated probability of mortality. The main factor going in an unexpected direction is obesity, but this effect may be overwhelmed if obese individuals are in poor or terrible health.

 While illuminating, these results are not directly usable in the full dynamic health and retirement model. The problem is that there are 8 medical conditions which can occur in any combination, and this means there are 256 different combinations. Since the medical conditions are carried as state variables in the dynamic model, this would increase the state space by a factor of 256, which would make the model computationally infeasible. We therefore try to group the conditions to reflect their relative effect on mortality. That is, some conditions may have a very large effect on mortality, while others may have relatively little effect. By grouping, we can cut down on the number of state variables and restore computational feasibility.

 The coefficients in Appendix Table 2.1 provide the basis for the groupings. For instance, an individual with heart problems at age 55 has the βX term in the mortality equation increase by 0.333 – 0.301 × 0.1 × (55 – 65) + 1.023 × 0.001 × (55 – 65)2 = 0.74. Performing this same exercise at ages 65 and 75 yields adjustments to the βX term of 0.33 and 0.13. Summing these three numbers yields 1.20. Doing this same exercise for the full list of eight medical conditions yields the following results:

 Hypertension 0.32

 Diabetes 1.30

 Cancer 2.66

 Lung Disease 0.80

 Heart Problems 1.20

 Stroke 0.51

 Psychiatric Problems -0.27

 Arthritis -1.02

 In this list, it is clear that cancer in the preceding survey has by far the greatest effect on mortality. Hypertension, psychiatric problems, and arthritis all appear to have minimal effects on mortality. One might reasonably ask why a condition such as stroke does not have a higher ranking. The answer appears to be that these mortality figures are based on the situation as of the preceding wave of the survey. If the individual had not had a stroke as of the preceding wave but had a stroke and died between waves, he would not be included in the stroke variable for this estimation. The stroke variable would include individuals who had had a stroke as of the previous wave but who were alive at the time of that wave, and it is reasonable that if the individual survived the stroke and lived to the previous wave, the mortality rate after that wave might not be as elevated as the total mortality rate for strokes. Cancer, on the other hand, generally does not kill immediately, so there is a much greater likelihood that someone who dies of cancer will have reported that cancer in the wave preceding the death.

 With this in mind, we will form three groups of conditions with regard to mortality. These are:

 Minimal Effect on Mortality Hypertension

 Psychiatric Problems

 Arthritis

 Moderate Effect on Mortality Diabetes

 Lung Disease

 Heart Problems

 Stroke

 Large Effect on Mortality Cancer

Individuals are assigned to the groups based on the most severe medical condition they have. For example, if an individual has both hypertension and heart problems, he would be assigned to the moderate group based on the heart problems. Individuals with no medical conditions are assigned to the first group, with minimal effect on mortality.

 Appendix Table 2.2 details estimates of the mortality hazard function using this grouping of conditions. The coefficients for the health categories and the personal characteristics are generally not very different from the estimation using the eight medical conditions separately. The coefficients for the mortality groups are in line with the coefficients of the medical conditions which comprise the groups in the estimation with the eight conditions.

 *B. Transitions Among Health States.*

 The transitions among health states are taken to depend on the previous health state, previous medical conditions, new medical conditions, and the personal characteristics. Given the previous health state, the probability of the current health state is taken to be an ordered probit. Let the health state in the previous year be ht-1, where the value of the health states are good, fair, poor, and terrible. For convenience, we can denote these states as 1, 2, 3, and 4. The probabilities of ht are then given by the equations:

 Pr (ht = 1) = F(bi1 + βiX)

 Pr (ht = 2) = F(bi2 + βiX) - F(bi1 + βiX)

 Pr (ht = 3) = F(bi3+ βiX) - F(bi2 + βiX)

 Pr (ht = 4) = 1 - F(bi3+ βiX)

where F is the cumulative normal distribution. The X variables include the previous medical conditions, new medical conditions, personal characteristics, and age. The coefficient vector βi depends on the health state in the previous year: i = ht-1.

 The four ordered probit equations (one for each of the health states in the previous year) must be estimated jointly because the health states are only observed every two years, and in the intermediate years the health state could be any of the four states. That is, for good health in year t-2 and fair health in year t, there are four possible sequences: good → good → fair, good → fair → fair, good → poor → fair, and good → terrible → fair. The estimation procedure allows for all of these pathways, but it requires that the four ordered probits be estimated jointly.

 The results of the estimation are presented in Appendix Table 2.3. Each panel of the table corresponds to a particular health state in the previous year. The coefficients labeled Break1, Break2, and Break3 in each panel are the bij’s in the above equations, where j runs from 1 to 3. The X vector includes a linear term and a quadratic term in age – 65. It also includes each of the previous medical conditions and new medical conditions and the personal characteristics, as well as these variables interacted with the linear and quadratic age terms. Negative coefficients of the X variables shift the probabilities for health transitions towards worse health. As expected, the coefficients for new medical conditions tend to be more negative than the coefficients for medical conditions which existed in the previous year.

 To illustrate, consider an individual in good health in the previous year who is 65, has none of the medical conditions, none of the negative personal characteristics, and is of average cognition. For such an individual, all of the X variables will be zero. The probability of staying in good health is F(1.214) = 0.888. the probability of transitioning to fair health is F(2.508) – F(1.214) = 0.106, the probability of transitioning to poor health is F(3.320) – F(2.508) = 0.006, and the probability of transitioning to terrible health is 1 – F(3.320) = 0.0005. Doing the same exercise for the different health states in the previous year gives the following table for the transitions:

 Health in Current Year

 Good Fair Poor Terrible

 Health Good 0.888 0.106 0.006 0.0005

 in Fair 0.321 0.615 0.063 0.001

Previous Poor 0.041 0.322 0.603 0.034

 Year Terrible 0.006 0.016 0.258 0.719

As mentioned, adding any of the medical conditions or negative personal characteristics will generally shift the probabilities towards the right side of this table.

 Using these estimates in the structural health and retirement model runs into the same problem as was encountered for the mortality estimates, namely, the number of combinations of the medical conditions is so large that it would be computationally infeasible to include them in the structural model. A similar solution to this problem is applied here, to group the medical conditions according to the degree to which they make transitions to a worse health state more likely. This is done by comparing the average negative shift of the βX term over the four categories of health in the previous year for both existing conditions and new conditions for the same three ages that were used in the mortality analysis.

 To illustrate, consider hypertension. For individuals initially in good health, the coefficients for a condition existing the previous year are -0.126, 0.054, and 0.311 for the main effect, the linear age term, and the quadratic age term, respectively. At age 55, the total effect of hypertension in the previous year for someone in good health that year is to change the βX term by -0.126 + 0.054 × 0.1 × (55 – 65) + 0.311 × 0.001 × (55 – 65)2 = -0.15. Similar calculations for ages 65 and 75 yield -0.13 and -0.04. Summing the values for these three ages gives -0.32. Doing analogous calculations for those in fair, poor, and terrible health in the previous year gives values of -0.15, -0.29, and 0.03. Using the coefficients for new episodes of hypertension gives values of -0.38, 0.00, -0.57, and 0.43 for those in good, fair, poor and terrible health in the previous year. Adding up the eight values (four for existing hypertension in the previous year and four for new hypertension) yields a value of -1.25.

 Doing the same exercise for all eight of the conditions yields the following results:

 Hypertension -1.25

 Diabetes -2.18

 Cancer -3.62

 Lung Disease -6.41

 Heart Problems -3.33

 Stroke -8.26

 Psychiatric Problems -6.21

 Arthritis -4.85

Hypertension appears to have the least effect on health transitions, perhaps because existing medications can control this disease fairly well, while stroke has the most effect on health for those who survive it. Among the remaining conditions, diabetes, cancer, and heart problems have relatively mild effects on the health transitions while lung disease, psychiatric problems and arthritis have more noticeable effects. This yields the following four groupings:

 Little or No Effect on Health Transitions: Hypertension

 Mild Effect on Health Transitions: Diabetes

 Cancer

 Heart Problems

 Moderate Effect on Health Transitions: Lung Disease

 Psychiatric Problems

 Arthritis

 Large Effect on Health Transitions: Stroke

Individuals with multiple medical conditions are assigned to the group with the largest effect of any of the individual conditions. Those with no medical conditions are assigned to the first group, with little or no effect on health transitions.

 Appendix Table 2.4 presents estimates of the health transitions based on these four groups of conditions. The coefficients for the personal characteristics and the break points are generally fairly similar to the corresponding coefficients estimated with the full set of conditions. The pattern of coefficients for the medical condition groups also follows the pattern that would be expected.

*C. Transitions Among Medical Condition Groups.*

 The previous analysis of mortality has grouped the medical conditions into three groups related to their effect on mortality, and the previous analysis of health transitions has grouped the medical conditions into four groups on the basis of their effects on health transitions. We may view the related groupings in a two-way chart as follows:

 Effect on Mortality

 Little or None Moderate Large

 ----------------------------------------------------------------------

 Little or None Hypertension

 Effect Mild Diabetes Cancer

 On Heart Problems

 Health

Transitions Moderate Arthritis Lung Disease

 Psychiatric Problems

 Large Stroke

 Since an individual can have more than one condition, and since the effect on mortality and health transitions is presumed to be at the greatest level of any of the individual conditions, some of the cells of this table may be occupied by combinations of the conditions. For instance, someone with diabetes and arthritis would be considered to be in the cell with a moderate effect on mortality (due to the diabetes) and a moderate effect on health transitions (due to the arthritis). Cells that are not occupied by individual conditions may nevertheless be occupied by combinations of the conditions. For example, someone with cancer and arthritis would be in the cell with a large effect on mortality and a moderate effect on health transitions. Taking these combinations into account, eight of the twelve potential cells in this table are occupied. They may be labeled as follows:

 Effect on Mortality

 Little or None Moderate Large

 --------------------------------------------------------------------

 Little or None C1 o o

 Effect

 On Mild o C2 C3

 Health

Transitions Moderate C4 C5 C6

 Large o C7 C8

Cells marked with a o do not correspond to any combination of conditions. By focusing on these eight combinations of conditions, rather than all 256 permutations of the eight conditions, we are able to provide reasonable variation in the ability of the conditions to affect mortality and the health transition probabilities while maintaining a number of states that makes the structural health and retirement computationally feasible.

 Transitions among these eight combinations is constrained by the fact that the questions for the conditions are of the form “Has a doctor ever told you that you have ….” Once this question is answered in the affirmative, it cannot logically be answered negatively in future years. This means that transitions up or to the left in the table above are logically not possible. Transitions occur when the respondent reports a new condition that has either a greater effect on mortality or a greater effect on health transitions than his previous combination of conditions. All of the combinations except C6 and C8 can be the result of single condition. For instance, an individual who previously had no conditions but now has lung disease would transition from C1 to C5. A transition to C5 might also result if an individual who previously reported diabetes now reports arthritis. This would be a transition from C2 to C5. A transition to C6 can occur if an individual who previously reported cancer now also reports lung disease (a transition from C3 to C6) or an individual who previously reported arthritis now also reports cancer (a transition from C4 to C6). Instances where individuals report two new conditions are quite rare, so we abstract from them and assume that, for instance, a transition directly from C1 to C8 does not occur because it would require both a new report of cancer and a new report of stroke. It is possible, however, for a transition from C1 to C3 to occur in one year due to cancer and from C3 to C8 to occur in a later year due to a stroke.

 Transitions to the various condition combinations are taken to depend on previous conditions, health status, and the personal characteristics. Each transition equation is taken to be the transition to a particular cell, and the sample for that equation is the group of individuals who can reach that cell with one new condition. For instance, the equation for transitions to cell C3 has as its sample the individuals who were in either cell C1 or C2 the previous period. Individuals who were in cells C4, C5, C6, C7, and C8 already have conditions which have a greater effect on health transitions than C3.

 Appendix Table 2.5 presents the results of the transition equations into the groups. The transition equations are of the form:

h = 0.01 \* exp(βX)

where h is the hazard for the transition and X is a vector of explanatory variables including previous conditions, health status, and personal characteristics. There is no equation for transition into C1, since this combination has the lowest effects both on mortality and on the probability of health transitions. The probability for remaining in C1 is simply the probability that none of the transitions out of C1 occur. The probability of transitioning into C2 is given by the first set of results in this table. Note that this equation is effectively the probability of developing diabetes or heart problems, given either no prior conditions or prior hypertension. The first row is the coefficients associated with a transition out of C1. Since the only combination that can transition into C2 is C1, everyone in the sample for this equation was previously in C1. Thus this row effectively contains the constant, age, and age squared terms for βX in the hazard formula. The other rows of this section of the table contain the coefficients for the various personal characteristics and health status variables, and their interactions with age and age squared. Note that there are no rows for average cognition and good health, which are the omitted categories. For the other personal characteristics, the omitted category is effectively not having the characteristic.

 The next section of the table is the probability of transitioning into C3 from either C1 or C2. This is effectively the probability of developing cancer, given either no prior conditions, prior hypertension, prior diabetes, or prior heart problems. The first two lines of results are the constant, age, and age squared terms depending on whether the prior combination is C1 (no prior conditions or only prior hypertension) or C2 (prior diabetes and/or prior heart problems). The remaining lines relate to the personal characteristics and prior health status and their interactions with age and age squared. The fact that the coefficients in the lines relating to C1 and C2 are approximately the same means that the probability of transitioning into C3 (by developing cancer) is about the same whether the individual is initially in C1 (no prior conditions on only prior hypertension) or C2 (prior diabetes and/or prior heart problems). This makes reasonable sense, since a diagnosis of cancer is likely to be relatively unrelated to hypertension, diabetes, or heart problems. Also note that the coefficient of the main effect for C1 is higher in the transition equation into C2 than it is for the transition equation into C3. This means that the probability of transitioning out of C1 is higher for a transition into C2 (diabetes or heart problems) than for a transition into C3 (cancer). This also makes reasonable sense, since the combination of diabetes and/or heart problems is probably more common than cancer.

 The remainder of Appendix Table 2.5 gives the coefficients for the transitions into the C4 through C8 combinations. In general, the results are much as would be expected. For instance, the strongest coefficient for obesity is with the transition into combination C2, which encompasses diabetes and heart problems. Since diabetes is the condition most frequently associated with obesity, this result is much of what one would anticipate. For many of the conditions, the transition probabilities increase significantly as the prior health state becomes worse, again what one might expect.

*D. Dynamic Simulation of the Health Model.*

 The mortality, health transition, and condition transition equation may be combined to form a dynamic model of health, where the results for one year become the explanatory variables for the next year. The dynamic framework we use may be sketched as follows:

1. The health status and condition combination at time t-1 determines mortality between t-1 and t.

2. For those who survive, the health status and condition combination at time t-1 determines the condition transition probabilities to the condition combination at time t.

3. The health status and condition combination at time t-1, along with the condition combination at time t, determines the health transition probabilities to the health status at time t. This allows, for instance, a new instance of heart problems at time t to have a different impact on health transitions than heart problems which already existed at t–1.

4. The new health status and condition combination at time t, as determined in steps 2 and 3, becomes the basis for the future changes at times t+1 and times future to t+1.

 We pay particular attention to making sure that the dynamic simulation of the health sub-model tracks reasonably well the population statistics on survival as well as the incidence of the various health states and condition combinations at various ages. As mentioned before, we already suspect that the mortality estimates generated by the HRS data may underestimate the aggregate mortality as revealed in the Social Security life tables. It turns out that this same issue arises with the conditions and health status. The frequency of transitions into the various medical conditions over a period of time often does not match the increase in the incidence of those conditions for those of different ages.

 To take one egregious example of this mismatch between the condition transitions and the incidence of conditions at various ages, consider hypertension. For those in the 50-59 age range the observed frequency in the HRS is 41.9%, and in the 60-69 age range it is 55.3%. Looking at the transitions of those who have not previously had hypertension in one wave to reporting hypertension in the wave two years later, the probabilities are 9.1% for 50-59 year olds and 9.4% for 60-69 year olds. Taking the average transition rate of 9.25% and dividing it by 2 to put it on an annual basis yields 4.625% per year. Starting at 41.9% figure for hypertension, the percentage without hypertension that year is 58.1%. If 4.625% of those develop hypertension over the course of the year, the next year the percentage with hypertension will be 44.6%. Repeating this process over a 10 year period ultimately yields 63.8% with hypertension at the end of the 10 year period. But the increase in observed incidence in hypertension only rose from 41.9% to 55.3% over the 10 year period. That is, there appears to be a substantial inconsistency between the incidence rates for new hypertension, which were used to estimate the transition equations, and the frequency of observed hypertension at different ages.

 One possible explanation might be that individuals with hypertension have higher mortality rates than those without hypertension. This is true, but the difference is not nearly enough to account for the inconsistency. Factoring in the differential mortality reduces the 63.8% figure at the end of 10 years to 62.2%, still substantially above the observed 55.3% figure for the cohort 10 years older. Another explanation is that individuals fail to report episodes of hypertension in earlier waves but something occurs which causes them to remember them in later waves. Thus, some of the reported new cases of hypertension may not be new after all. The problem with this explanation is that even if observations for the first two year period observed for a particular individual are discarded, the transition percentages do not move by very much.

 Whatever the causes of the inconsistency between the transition probabilities and the observed frequencies, it is clear that using the estimated transition probabilities unadjusted in a dynamic simulation, where the results of the transitions at one age form the frequencies for the next age, will result in an increasing divergence between the simulated mortality, condition frequencies, and health status distribution relative to the observed population frequencies for the same quantities as the individual ages. Since it is the probabilities of survival, the probabilities of the medical conditions, and the probabilities of the health states that are important for the retirement model, it is critical that the transition probabilities be adjusted so that the dynamic simulation yields sensible results. We do this by adjusting the constants in the various estimated transition equations up or down. This will allow the dynamic simulations to track the observed frequencies fairly well while maintaining the differential effects of the conditions and health state on the various transition frequencies.

 The nature of the adjustment is different for mortality, for medical conditions, and for the health states. We have detailed population statistics on mortality, so for mortality we can use a proportional adjustment factor which is a sixth-degree polynomial in age. For transitions into the medical condition groups, we use a simple proportional adjustment factor which adjusts the hazard rate of each group up or down. The adjustment factors are allowed to be different for each group. For transitions among the health states, we adjust transitions out of the current state and into the other states proportionately. The adjustment factors are allowed to be different depending on the current state.

 We choose these adjustment factors, along with the initial percentages of the condition groups and the health states at age 50, so as to minimize the difference between the simulated results and the observed frequencies. For survival statistics, the component of the quantity to be minimized is:



where a is age and the S’s are the sample survival probabilities from age 50 for the simulated sample and for the Social Security survival tables. HRS weights are used for these probabilities in order to make the sample survival probabilities for the simulated sample comparable to the Social Security survival tables. For the condition probabilities, the component of the quantity to be minimized is



where i refers to the eight condition groups, j refers the five 10 year age ranges 50-59, 60-69, 70-79, 80-89, and 90-99, and C refers to the average percentage of the sample in the condition group i and age range j for the simulated sample and for the observed sample. For the health states, the component of the quantity to be minimized is



where i refers to the four health states, j again refers to the same five 10 year age ranges, and H refers to the average percentage in the given health state and age range for the simulated sample and for the observed sample. For the condition groups and health states, we use the 10 year averages to reduce the year-to-year fluctuations in the observed data, which become somewhat thin when used in single year age ranges.

 The overall quantity to be minimized is simple the sum of these three components

Q = Qs + Qc + Qh

Since each of the components is simply the average squared deviation of the simulated results from the observed results over the number of cells in the average, this overall quantity gives equal weight to fitting the survival probabilities, the condition group probabilities, and the health state probabilities. Q is minimized with respect to the adjustment factors for the estimated transition equations and the initial values at age 50 of probabilities of the condition groups and health states. For the condition groups, the adjustment factors for transitions into the indicated conditions are:

 C2 0.853

 C3 0.561

 C4 0.529

 C5 0.494

 C6 0.489

 C7 0.798

 C8 0.597

There are no values for C1 because individuals can transition out of C1, but not into it. Values less than unity indicate that the transitions into the conditions, as calculated from the estimated transition equations, are proportionately reduced in order to make the observed frequencies of the conditions approximately equal to the observed frequencies in the dynamic simulation. This is what one would expect based on the previous discussion regarding hypertension. For the health transitions, the adjustment factors for the equations out of specific health states are as follows:

 Good 0.932

 Fair 0.934

 Poor 0.911

 Terrible 1.037

These figures indicate that the health transition equations do not need substantial modification in order to make the simulated results track the observed results reasonably well.

 Comparisons of the simulated values for the sample to the observed values are given in Appendix Tables 2.6, 2.7, and 2.8. Appendix Table 2.6 compares the survival probabilities for the sample, given survival to age 50, for the simulation and for the Social Security life tables. Values were computed for each age, but only values at 5 year intervals are given to reduce clutter. At some ages the simulated survival rates exceed the observed survival rates and at other ages the reverse is true. The absolute difference never exceeds 0.8 percentage points, however, even for the ages not included in the table, and the average arithmetic difference is on the order of 0.1 percentage points. All in all, this should give a reasonable approximation to life expectancies, particularly when two simulations are compared.

 Appendix Table 2.7 compares the probabilities of being in the various condition groups, by 10 year age intervals, for the simulation and for the frequencies computed from the sample. In general, the simulation does a fairly good job of tracking the changes in the observed frequencies. The average absolute deviation between the simulation and the observed frequencies is around one percentage point. The largest discrepancies occur in the oldest age range, but the number of observations in this range is relatively low, with a confidence range of approximately plus or minus two percentage points. The simulation captures the trend of all of the condition groups, and it frequently captures turning points in the frequencies, where the frequency is increasing at earlier ages but declines at later ages, as well.

 Appendix Table 2.8 compares the probabilities of being in the various health states, again by 10 year age intervals, for the simulation and for the frequencies computed from the sample. The simulation does an even better job in tracking the health states than in tracking the condition groups, with the average absolute deviation between the simulated and observed health states less than 0.7 percentage points. The major trends in health are tracked quite well, and the simulation succeeds in reflecting the peak of fair health, which rises in the early age ranges as people fall from good health to fair health and then falls in the later age ranges as people fall from fair health to poor or terrible health.

Appendix Table 2.1

Mortality Estimates with Full Set of Conditions

 Interaction Terms

 Main Effect × 0.1 × (age – 65) × 0.001 × (age – 65)2

Base -0.345 \*\*\* 1.002 \*\*\* -0.166

Medical Conditions

 Hypertension 0.078 -0.158 \*\* 0.446 \*

 Diabetes 0.416 \*\*\* -0.181 \*\*\* 0.244

 Cancer 0.820 \*\*\* -0.507 \*\*\* 0.992 \*\*\*

 Lung Disease 0.290 \*\*\* 0.059 -0.370

 Heart Problems 0.333 \*\*\* -0.301 \*\*\* 1.023 \*\*\*

 Stroke 0.146 \*\* -0.117 0.343

 Psychiatric Problems -0.061 0.108 -0.450

 Arthritis -0.371 \*\*\* -0.090 0.458\*

Personal Characteristics

 Smoking 0.435 \*\*\* -0.073 0.006

 Drinking 0.019 0.020 -0.468

 Obesity -0.470 \*\*\* 0.101 -0.332

 Depression 0.131 \*\* -0.037 -0.327

 Cognition – Impaired 0.270 \*\*\* 0.012 -0.082

 Cognition - Excellent -0.163 \*\* 0.022 -0.670

Health Status

 Fair 0.630 \*\*\* -0.186 \* 0.227

 Poor 1.307 \*\*\* -0.039 -0.587

 Terrible 1.834 \*\*\* 0.061 -0.965 \*

Significance levels: \* 90%, \*\* 95%, and \*\*\* 99%

Appendix Table 2.2

Mortality Estimates with Grouped Conditions

 Interaction Terms

 Main Effect × 0.1 × (age – 65) × 0.001 × (age – 65)2

Base -0.608 \*\*\* 0.951 \*\*\* 0.166

Mortality Group

 Moderate 0.693 \*\*\* -0.395 \*\*\* 0.933 \*\*\*

 High 1.314 \*\*\* -0.736 \*\*\* 1.376 \*\*\*

Personal Characteristics

 Smoking 0.454 \*\*\* -0.020 -0.150

 Drinking 0.011 0.028 -0.484

 Obesity -0.435 \*\*\* 0.037 -0.229

 Depression 0.157 \*\*\* -0.021 -0.488 \*

 Cognition – Impaired 0.276 \*\*\* 0.013 -0.101

 Cognition - Excellent -0.161 \*\* 0.040 -0.758

Health Status

 Fair 0.534 \*\*\* -0.197\*\* 0.353

 Poor 1.109 \*\*\* -0.073 -0.399

 Terrible 1.771 \*\*\* 0.017 -0.810

Significance levels: \* 90%, \*\* 95%, and \*\*\* 99%

Appendix Table 2.3

Health Transition Estimates with Full Set of Conditions

 Interaction Terms

 Main Effect × 0.1 × (age – 65) × 0.001 × (age – 65)2

 A. Previous Health State: Good

Break 1 1.214 \*\*\*

Break 2 2.508 \*\*\*

Break 3 3.320 \*\*\*

Base -0.187 \*\*\* -0.357 \*

Existing Conditions

 Hypertension -0.126 \*\*\* 0.054 \*\* 0.311

 Diabetes -0.161 \*\*\* 0.059 -0.398

 Cancer -0.144 \*\*\* 0.112 \*\* -0.039

 Lung Disease -0.291 \*\*\* 0.088 0.363

 Heart Problems -0.205 \*\*\* -0.046 0.263

 Stroke -0.201 \*\*\* 0.047 0.434

 Psychiatric Problems -0.177 \*\*\* 0.035 -0.167

 Arthritis -0.364 \*\*\* 0.068 \*\*\* -0.104

New Conditions

 Hypertension -0.079 0.115 \*\* -0.736

 Diabetes -0.267 \*\*\* -0.020 1.133

 Cancer -0.336 \*\*\* -0.016 0.129

 Lung Disease -0.652 \*\*\* 0.106 0.746

 Heart Problems -0.269 \*\*\* 0.107 \* -0.924 \*

 Stroke -0.603 \*\*\* 0.289 \*\* -2.685 \*\*\*

 Psychiatric Problems -0.246 \*\* 0.062 -1.550 \*\*

 Arthritis -0.403 \*\*\* 0.208 \*\*\* -1.086 \*\*\*

Personal Characteristics

 Smoking -0.143 \*\*\* 0.022 0.370

 Drinking -0.016 -0.052 -0.456

 Obesity -0.172 \*\*\* 0.033 -0.296

 Depression -0.215 \*\*\* 0.062 0.090

 Cognition - Impaired -0.150 \*\*\* -0.047 -0.865 \*

 Cognition – Excellent 0.138 \*\*\* -0.057 \*\* 0.283

Appendix Table 2.3 (cont.)

 Interaction Terms

 Main Effect × 0.1 × (age – 65) × 0.001 × (age – 65)2

 B. Previous Health State: Fair

Break 1 -0.466 \*\*\*

Break 2 1.521 \*\*\*

Break 3 3.028 \*\*\*

Base -0.186 \*\*\* -0.128

Existing Conditions

 Hypertension -0.034 0.014 -0.242

 Diabetes -0.157 \*\*\* 0.000 0.207

 Cancer -0.070 \*\* 0.053 -0.250

 Lung Disease -0.275 \*\*\* -0.072 \* 0.716 \*\*

 Heart Problems -0.122 \*\*\* 0.055 \* -0.072

 Stroke -0.151 \*\*\* 0.020 -0.009

 Psychiatric Problems -0.110 \*\*\* 0.064 \* -0.695 \*\*

 Arthritis -0.254 \*\*\* 0.060 \*\* -0.372 \*\*

New Conditions

 Hypertension -0.011 0.027 0.168

 Diabetes -0.068 0.092 -0.768

 Cancer -0.205 \*\*\* 0.201 \*\*\* -0.995 \*

 Lung Disease -0.300 \*\*\* -0.096 -0.125

 Heart Problems -0.270 \*\*\* 0.004 0.031

 Stroke -0.477 \*\*\* -0.108 0.559

 Psychiatric Problems -0.381 \*\*\* -0.143\* 0.156

 Arthritis -0.324 \*\*\* 0.013 0.213

Personal Characteristics

 Smoking -0.146 \*\*\* -0.015 0.075

 Drinking 0.007 -0.034 0.212

 Obesity -0.159 \*\*\* -0.032 0.064

 Depression -0.223 \*\*\* 0.066 \* 0.047

 Cognition - Impaired -0.194 \*\*\* 0.022 0.417

 Cognition - Excellent 0.085 \*\*\* 0.041 -0.126

Appendix Table 2.3 (cont.)

 Interaction Terms

 Main Effect × 0.1 × (age – 65) × 0.001 × (age – 65)2

 C. Previous Health State: Poor

Break 1 -1.738 \*\*\*

Break 2 -0.350 \*\*\*

Break 3 1.827 \*\*\*

Base -0.065 -0.837 \*\*\*

Existing Conditions

 Hypertension -0.096 \*\*\* 0.044 -0.029

 Diabetes -0.133 \*\*\* 0.045 -0.243

 Cancer -0.055 0.034 -0.054

 Lung Disease -0.148 \*\*\* 0.010 -0.419

 Heart Problems -0.081 \*\*\* -0.017 0.062

 Stroke -0.264 \*\*\* -0.023 0.257

 Psychiatric Problems -0.129 \*\*\* -0.080 \*\* 0.441

 Arthritis -0.199 \*\*\* 0.071 \*\* 0.212

New Conditions

 Hypertension -0.193 \*\*\* -0.031 0.030

 Diabetes -0.040 -0.097 -0.374

 Cancer -0.196 \*\*\* 0.032 -0.436

 Lung Disease -0.238 \*\*\* -0.012 -0.052

 Heart Problems -0.171 \*\*\* 0.008 0.115

 Stroke -0.611 \*\*\* 0.110 -0.390

 Psychiatric Problems -0.376 \*\*\* -0.226 \*\*\* 0.453

 Arthritis -0.233 \*\*\* 0.052 -0.465

Personal Characteristics

 Smoking -0.163 \*\*\* -0.037 1.146 \*\*\*

 Drinking 0.018 -0.015 0.055

 Obesity -0.034 -0.063 \* -0.250

 Depression -0.116 \*\*\* 0.042 -0.293

 Cognition – Impaired -0.101 \*\* 0.044 -0.828 \*\*\*

 Cognition – Excellent 0.117 \*\*\* 0.011 0.051

Appendix Table 2.3 (cont.)

 Interaction Terms

 Main Effect × 0.1 × (age – 65) × 0.001 × (age – 65)2

 D. Previous Health State: Terrible

Break 1 -2.485 \*\*\*

Break 2 -2.003 \*\*\*

Break 3 -0.580 \*\*\*

Base -0.134 -0.136

Existing Conditions

 Hypertension 0.035 0.068 -0.355

 Diabetes -0.013 -0.031 -0.005

 Cancer -0.057 0.206 \*\* -0.710

 Lung Disease -0.213 \*\*\* 0.003 0.541

 Heart Problems 0.021 0.091 -0.670

 Stroke -0.258 \*\*\* -0.213 \*\*\* 1.191 \*\*

 Psychiatric Problems -0.219 \*\*\* 0.114 \* -0.316

 Arthritis 0.066 -0.021 0.101

New Conditions

 Hypertension 0.133 -0.020 0.151

 Diabetes 0.207 0.058 -0.955

 Cancer 0.132 0.047 -1.769

 Lung Disease -0.239 -0.519 \*\*\* 1.531

 Heart Problems 0.044 -0.245 \* 0.360

 Stroke -0.176 -0.313 0.436

 Psychiatric Problems -0.282 \* 0.237 -0.584

 Arthritis 0.266 \* 0.078 -1.049

Personal Characteristics

 Smoking 0.027 0.015 -0.901

 Drinking 0.272 \*\*\* -0.035 -0.929

 Obesity 0.025 -0.002 -0.310

 Depression -0.168 \*\*\* 0.077 0.088

 Cognition - Impaired -0.241 \*\*\* -0.176 \*\* 1.413 \*\*\*

 Cognition - Excellent 0.133 0.041 -0.272

Appendix Table 2.4

Health Transition Estimates with Grouped Conditions

 Interaction Terms

 Main Effect × 0.1 × (age – 65) × 0.001 × (age – 65)2

 A. Previous Health State: Good

Break1 1.175 \*\*\*

Break2 2.450 \*\*\*

Break3 3.242 \*\*\*

Base -0.153 \*\*\* -0.173

Existing Conditions

 Mild Effect -0.241 \*\*\* -0.023 0.034

 Moderate Effect -0.508 \*\*\* 0.021 0.076

 Large Effect -0.561 \*\*\* 0.008 0.434

New Conditions

 Mild Effect -0.320 \*\*\* 0.011 -0.622

 Moderate Effect -0.521 \*\*\* 0.182 \*\*\* -0.543

 Large Effect -0.784 \*\*\* 0.354 \*\*\* -2.638 \*\*\*

Personal Characteristics

 Smoking -0.120 \*\*\* 0.017 0.170

 Drinking -0.005 -0.035 -0.478

 Obesity -0.205 \*\*\* 0.045 -0.200

 Depression -0.230 \*\*\* 0.032 0.199

 Cognition - Impaired -0.155 \*\*\* -0.046 -0.762 \*

 Cognition – Excellent 0.144 \*\*\* -0.058 \*\* 0.272

 B. Previous Health State: Fair

Break1 -0.453 \*\*\*

Break2 1.521 \*\*\*

Break3 2.985 \*\*\*

Base -0.156 \*\*\* -0.129

Existing Conditions

 Mild Effect -0.200 \*\*\* -0.001 -0.053

 Moderate Effect -0.412 \*\*\* 0.020 -0.363

 Large Effect -0.529 \*\*\* 0.038 -0.298

Appendix Table 2.4 (cont.)

 Interaction Terms

 Main Effect × 0.1 × (age – 65) × 0.001 × (age – 65)2

 B. Previous Health State: Fair (cont.)

New Conditions

 Mild Effect -0.097 0.066 -0.854

 Moderate Effect -0.341 \*\*\* -0.007 -0.236

 Large Effect -0.557 \*\*\* -0.113 0.545

Personal Characteristics

 Smoking -0.171 \*\*\* -0.017 0.185

 Drinking 0.014 -0.037 0.206

 Obesity -0.188 \*\*\* -0.028 0.130

 Depression -0.247 \*\*\* 0.057 \* 0.062

 Cognition - Impaired -0.189 \*\*\* 0.029 0.163

 Cognition - Excellent 0.082 \*\*\* 0.044 \* -0.081

 C. Previous Health State: Poor

Break1 -1.639 \*\*\*

Break2 -0.245 \*\*\*

Break3 1.915 \*\*\*

Base -0.109 \* -1.697 \*\*\*

Existing Conditions

 Mild Effect -0.284 \*\*\* 0.191 \*\*\* 0.540

 Moderate Effect -0.524 \*\*\* 0.072 1.332 \*\*\*

 Large Effect -0.763 \*\*\* 0.035 1.440 \*\*\*

New Conditions

 Mild Effect -0.036 0.141 -2.214

 Moderate Effect -0.209 \*\* 0.090 -1.094 \*

 Large Effect -0.671 \*\*\* 0.126 -0.447

Personal Characteristics

 Smoking -0.175 \*\*\* -0.047 1.095 \*\*\*

 Drinking 0.033 0.009 -0.023

 Obesity -0.056 \* -0.049 -0.302

 Depression -0.169 \*\*\* 0.031 -0.067

 Cognition - Impaired -0.096 \*\* 0.055 -0.879 \*\*\*

 Cognition - Excellent 0.114 \*\*\* 0.012 0.189

Appendix Table 2.4 (cont.)

 Interaction Terms

 Main Effect × 0.1 × (age – 65) × 0.001 × (age – 65)2

 D. Previous Health State: Terrible

Break1 -2.787 \*\*\*

Break2 -2.336 \*\*\*

Break3 -0.901 \*\*\*

Base 0.047 0.421

Existing Conditions

 Mild Effect 0.745 \*\*\* -0.153 -2.314 \*

 Moderate Effect 0.294 \* -0.133 -0.730

 Large Effect 0.028 -0.287 \* 0.124

New Conditions

 Mild Effect 0.352 - -

 Moderate Effect -0.167 -0.220 1.578

 Large Effect -0.261 \* -0.241 0.285

Personal Characteristics

 Smoking -0.009 -0.009 -0.527

 Drinking 0.211 \*\* -0.091 -0.264

 Obesity 0.029 0.006 -0.227

 Depression -0.208 \*\*\* 0.152 \*\*\* -0.387

 Cognition - Impaired -0.230 \*\*\* -0.193 \*\*\* 1.357 \*\*\*

 Cognition – Excellent 0.135 0.066 -0.332

Appendix Table 2.5

Transitions Among the Grouped Conditions

 Interaction Terms

 Main Effect × 0.1 × (age – 65) × 0.001 × (age – 65)2

Transitions to Group 2

From

 Group 1 0.773 \*\*\* 0.398 \*\*\* -0.441

Personal Characteristics

 Smoking 0.077 -0.140 0.236

 Drinking -0.038 0.021 -0.453

 Obesity 0.577 \*\*\* -0.165 0.160

 Depression 0.298 0.048 -0.645

 Cognition - Impaired -0.500 \* -0.349 3.491 \*\*

 Cognition - Excellent 0.058 0.139 -0.024

Health Status

 Fair 0.070 -0.169 0.246

 Poor -0.112 -0.264 1.560

 Terrible 0.085 -0.031 0.859

Transitions to Group 3

From

 Group 1 0.240 \*\* 0.750 \*\*\* -2.539 \*\*

 Group 2 0.268 0.688 \*\*\* -2.333

Personal Characteristics

 Smoking 0.201 0.084 0.386

 Drinking -0.005 -0.098 3.064 \*

 Obesity -0.103 -0.262 -0.076

 Depression -0.011 -0.375 2.562

 Cognition - Impaired -0.649 -0.124 -1.600

 Cognition - Excellent 0.216 0.105 -1.698

Health Status

 Fair 0.169 0.086 -2.300

 Poor -1.271 \* 1.616 -5.002

 Terrible -0.844 0.027 2.140

Appendix Table 2.5 (cont.)

 Interaction Terms

 Main Effect × 0.1 × (age – 65) × 0.001 × (age – 65)2

Transitions to Group 4

From

 Group 1 1.370 \*\*\* 0.046 -0.469

Personal Characteristics

 Smoking 0.031 -0.123 -1.725

 Drinking 0.038 0.065 0.251

 Obesity 0.225 \*\* -0.002 0.341

 Depression 0.092 -0.162 -0.908

 Cognition - Impaired 0.129 -0.171 0.826

 Cognition - Excellent -0.112 -0.036 0.881

Health Status

 Fair 0.378 \*\*\* 0.107 0.681

 Poor 0.727 \*\*\* 0.044 1.530

 Terrible 0.679 \* -0.427 0.730

Transitions to Group 5

From

 Group 1 -0.128 0.503 \*\*\* -1.034

 Group 2 1.549 \*\*\* 0.179 \*\* -0.322

 Group 4 1.146 \*\*\* 0.352 \*\*\* -0.396

Personal Characteristics

 Smoking 0.324 \*\*\* 0.004 0.588

 Drinking -0.013 -0.149 \* 0.953

 Obesity 0.354 \*\*\* -0.105 -0.510

 Depression 0.052 -0.697

 Cognition – Impaired 0.006 0.008 0.209

 Cognition - Excellent -0.153 \*\* 0.058 -0.121

Health Status

 Fair 0.406 \*\*\* -0.081 0.284

 Poor 0.619 \*\*\* -0.166 \* -0.005

 Terrible 0.748 \*\*\* -0.014 -0.213

Appendix Table 2.5 (cont.)

 Interaction Terms

 Main Effect × 0.1 × (age – 65) × 0.001 × (age – 65)2

Transitions to Group 6

From

 Group 3 1.909 \*\*\* 0.083 0.183

 Group 4 0.390 \*\*\* 0.696 \*\*\* -2.634 \*\*\*

 Group 5 0.257 \*\* 0.698 \*\*\* -2.218 \*\*

Personal Characteristics

 Smoking 0.136 -0.242 \* -2.075

 Drinking 0.068 0.282 -1.282

 Obesity 0.086 0.079 -0.046

 Depression 0.085 -0.095 -0.123

 Cognition - Impaired 0.008 -0.081 0.946

 Cognition - Excellent 0.115 -0.176 1.007

Health Status

 Fair 0.176 \* -0.173 0.709

 Poor 0.492 \*\*\* -0.148 -0.483

 Terrible 0.020 -0.292 1.281

Transitions to Group 7

From

 Group 1 -0.806 \*\*\* 0.880 \*\*\* -2.268

 Group 2 -0.128 0.676 \*\*\* -2.088

 Group 4 -0.855 \*\*\* 0.667 \*\*\* 0.057

 Group 5 -0.429 \*\*\* 0.586 \*\*\* 0.259

Personal Characteristics

 Smoking 0.466 \*\*\* -0.421\*\*\* -0.252

 Drinking -0.404 \*\* -0.110 0.429

 Obesity -0.097 -0.394 \*\* 2.105 \*

 Depression 0.450 \*\*\* 0.127 -2.084 \*\*

 Cognition - Impaired 0.506 \*\*\* -0.162 0.298

 Cognition - Excellent -0.328 \*\* 0.241 -0.081

Health Status

 Fair 0.105 0.116 -1.056

 Poor 0.428 \*\*\* -0.047 -0.814

 Terrible 0.689 \*\*\* 0.327 -2.368

Appendix Table 2.5 (cont.)

 Interaction Terms

 Main Effect × 0.1 × (age – 65) × 0.001 × (age – 65)2

Transitions to Group 8

From

 Group 3 -0.444 0.422 1.456

 Group 6 -0.662 \* 0.061 4.007

 Group 7 -0.321 0.527 2.270

Personal Characteristics

 Smoking 0.496 \* 0.554 -3.998

 Drinking -0.470 -0.075 1.650

 Obesity 0.236 0.017 -4.647

 Depression 0.527 \*\* -0.309 -0.298

 Cognition - Impaired -0.170 0.406 -1.756

 Cognition - Excellent -0.475 0.974 -3.522

Health Status

 Fair 0.204 0.282 -4.284

 Poor 0.656 \*\* 0.000 -2.727

 Terrible 0.842 \*\* -0.259 -1.606

Appendix Table 2.6

Survival Rates

 Social Security

 Age Simulation Survival Tables

 50 100% 100%

 55 95.6 96.2

 60 90.8 91.1

 65 84.4 84.2

 70 75.1 74.3

 75 61.6 61.8

 80 45.8 46.3

 85 29.4 29.2

 90 14.5 14.3

 95 4.3 4.9

Appendix Table 2.7

Frequencies of Medical Condition Groups

Age Range 50-59 60-69 70-79 80-89 90-99

Condition Group Simulated

 1 42.4% 27.0% 15.8% 9.1% 5.4%

 2 10.1 10.1 8.9 7.8 7.3

 3 0.8 2.8 4.8 5.3 4.2

 4 22.8 24.2 20.9 16.9 15.2

 5 16.5 23.7 29.2 31.6 29.6

 6 4.0 6.4 11.2 16.2 20.4

 7 3.3 5.1 7.1 9.1 11.4

 8 0.1 0.7 2.1 4.1 6.5

 Observed

 1 44.2 24.6 13.4 9.3 8.2

 2 10.0 9.3 9.3 8.0 7.3

 3 1.7 3.1 4.6 4.9 5.3

 4 23.3 24.5 18.8 14.8 18.4

 5 15.1 25.7 29.7 31.0 28.9

 6 2.5 7.0 14.0 18.3 16.5

 7 3.0 5.1 8.1 10.0 10.0

 8 0.2 0.7 2.0 3.8 5.4

Appendix Table 2.8

Frequencies of Health States

Age Range 50-59 60-69 70-79 80-89 90-99

 Health State Simulated

 Good 54.6% 42.5% 30.3% 18.9% 11.0%

 Fair 26.9 32.5 35.1 33.2 27.1

 Poor 12.7 19.1 25.8 33.2 37.4

 Terrible 5.9 5.9 8.7 14.7 24.5

 Observed

 Good 54.5 43.3 29.5 18.6 11.9

 Fair 26.8 32.1 37.3 33.3 25.1

 Poor 12.8 18.0 24.6 32.9 39.0

 Terrible 5.9 6.6 8.6 15.2 24.1

Appendix 3:

 This appendix reports the equations used to incorporate variation in layoffs, SSDI, medical expenditures and nursing home expenditures into the model. Probit equations are estimated as described in the text. Where necessary, adjustments are made so that outcomes conform to population totals.

The probability of SSDI is estimated with a standard probit formulation. The results are given in Appendix Table 3.1. The equation for the transition into SSDI is subject to the same tracking issues that we faced with health status and medical conditions, namely, applying this equation over time may yield percentages of individuals on SSDI which start to diverge from the observed percentages. To correct this problem, we apply a constant multiplicative factor to the transition equation so that a dynamic simulation yields approximately the same results as are observed. Effectively, this means that the estimated equation determines the relative effects of health status and the condition groups on SSDI eligibility, and the factor raises or lowers the transition probabilities to give the correct SSDI percentages over time.

 The SSDI transition equation is adjusted to give transitions into SSDI, and the adjustment factor is chosen to minimize the squared differences between the simulated percentages on SSDI at various ages and the observed percentages. The results are given in Appendix Table 3.2.

*D. Medical Expenditures.*

 Another uncertainty concerns medical costs for those who are uninsured. Large medical costs may be fairly uncommon, but when they occur they have the potential to reduce sharply the accumulated retirement savings if the individual is uninsured.

The distribution of health care costs is estimated for a sample of individuals who reported that they did not have any health insurance in the year. The functional form was a log-normal distribution of the type

Prob (cost < x) = F[(ln(x) – (μ + Xβ)) / σ]

where F is the standard cumulative normal distribution and X contains the health status categories and the condition groups. The parameter estimates are given in Appendix Table 3.3. Having poor or terrible health raises medical expenses significantly, as does having any of the medical conditions. Among the medical conditions, being in group 4 (arthritis and/or psychological problems) raises costs the least and being in group 8 (cancer and stroke) raises them the most.

 Since most of the time the expenses are relatively small, the retirement model focuses on the larger expenditures which may effectively bankrupt the couple. It groups expenditures by percentile groups, with the groups divided by expenditures at the 50th, 80th, 90th, 95th, 97.5th, 99th, and 99.5th percentiles. The probability of being in each of these eight groups defined by these breakpoints is 50%, 30%. 10%, 5%, 2.5%, 1.5%, 0.5%, and 0.5%. The average expenditure within each of these percentile ranges depends on the health status and condition group of the individual. Appendix Table 3.4 gives the average expenditure for each percentile range for the various health status and condition groups in 1992 dollars, as calculated from the log-normal distribution and the parameters estimated in Appendix Table 3.3. The table reflects the possibility of very high expenditures with relatively low probabilities. The pattern of expenditures with respect to health status and condition groups follows the pattern that would be expected on the basis of the estimated parameters. Again, bear in mind that these are estimates for uninsured individuals below the Medicare eligibility age.

*E. Nursing Home Expenditures.*

 For HRS respondents who die between surveys, the study attempts to conduct an “exit” interview with someone who was familiar with the respondent’s situation, usually a close relative. Among the questions asked are whether the individual was residing at a nursing home at the time of death, and if so, how long he had been in the nursing home. The study was careful to distinguish between where the individual was residing at the time of death and the actual location at death, which is usually in a hospital. Appendix Table 3.5 gives the percentage of HRS respondents from all cohorts and all waves who resided in nursing homes at the time of their death, broken down by health status and condition group. Relatively few individuals in their 50’s reside in nursing homes at the time of their death, but the number rises rapidly with age, with three-fifths of individuals over 90 years old residing in nursing homes at the time of their death. The percentage in nursing homes rises with deteriorating health status, as one might expect. The variation among those in different medical condition groups is much less noticeable, although there is some suggestion that more severe conditions are associated with an increased probability of residing in a nursing home at the time of death.

 Appendix Table 3.6 gives the distribution of the lengths of stays before death among those who resided in nursing homes at the times of their deaths. The first part of the table gives the 50th, 75th, and 90th percentile lengths of stay by age range for those in the four health status categories. The second part of the table gives the corresponding figures for those in the eight condition groups. The patterns among the health status categories and condition groups largely mirrors the corresponding patterns in the probability of residing in nursing homes. Overall, the distribution of stays looks relatively stable during the first three decades shown in the table. The median stay is two or three months, rising to not quite a year at the 75th percentile and about three years at the 90th percentile. As individuals age into their 80’s and 90’s, the distribution shifts markedly to the right, so that by the 90’s the median stay is not quite a year, the 75th percentile is a little over two years, and the 90th percentile is over four years.

 Appendix Table 3.7 presents the results of a probit estimate for the probability that the individual was residing in a nursing home at the time of his death. Explanatory variables include a quadratic in age, health status, and the condition group. The most significant coefficients are the linear term in age and a health status that is “terrible.” These results should come as no surprise given the statistics reported in Appendix Table 3.5. Appendix Table 3.8 calculates the probabilities of nursing home residence, broken down by health status and condition group, for the relevant sample using the coefficients of the probit estimate. A comparison of these numbers with the corresponding observed percentages in Appendix Table 3.5 suggests that the fitted numbers have the same patterns and overall levels as the observed numbers, although the fitted numbers tend to be a bit smoother. Bear in mind that the number of deaths is fairly low for individuals in the younger age ranges and for those in good health in the older age ranges, so that the confidence intervals for the observed figures can be several percentage points wide.

 Those who are not residing in a nursing home at the time of death will not have expenses related to the final nursing home stay. For those who do reside in a nursing home at the time of death, the expenses will be proportional to the length of time that the individual has been in the nursing home. Using the observed distributions by age, health status, and condition group results in cells that are thin enough that the distributions are quite uneven, so we estimate a parametric distribution in order to smooth out the distributions. The main problem is that the observed distribution of length of stay has a large number of relatively short observations but a long tail of lengthier observations. The usual distributions, such as the log-normal or exponential, which reflect the mass of observations at short durations, tend to underestimate the tail of the longer durations.

 To address this issue, we formulate the duration problem as a two-part process. First, we use a probit equation to determine the probability that the duration will be shorter than two months. For the retirement model, we will end up grouping durations by groups, with the shortest interval being zero to two months. Thus the probit equation effectively determines the probability of being in the shortest interval. For durations longer than two months, we use an exponential distribution. The resulting contribution to the likelihood function for an individual observation is

 Li = F(Xβ) if z < 2

 Li = [1 – F(Xβ)] (Xγ) e –(Xγ) ( z – 2) if z ≥ 2

where Li is the contribution to the likelihood function, F is the cumulative standard normal distribution, z is the observed length of the final nursing home stay, and X is a vector of explanatory variables including a quadratic in age, health status, and the condition group. β is the parameter vector for the probit, and γ is the parameter vector for the exponential distribution. Note that the integral of the second term between 2 and infinity is simply 1 – F(Xβ), the probability that the stay will be longer than 2 months.

 Estimates of this distribution are given in Appendix Table 3.9. The top part of the table refers to the estimates β for the probit part of the distribution, and the bottom part refers to the estimates γ for the exponential part of the distribution. As with the estimates for the probability of a nursing home stay, the most significant coefficients for both the probit and the exponential parts of the distribution are the linear term in age and terrible health status. The quadratic in age is also significant in the exponential part of the distribution, though not in the probit. In general, the age and health coefficients are in the expected direction.

 Appendix Table 3.10 compares the observed distribution of nursing home durations with the fitted durations by age range. Again the fitted distribution preserves the overall pattern and levels of the observed distribution, although the correspondence is not as great as with the frequency of nursing home residence in the previous tables, especially at the younger ages. But bear in mind that there were only 2024 observations of deaths of individuals residing in a nursing home for whom data on length of stay is available. Relatively few of these occurred at younger ages, both because the mortality rates among younger ages is substantially less and because, conditional on mortality, younger individuals are much less likely to reside in a nursing home. Hence, the number of individuals in their 50’s in Appendix Table 36.10 is not much more than a couple of dozen, and the percentages in the various cells in the observed distribution have very wide confidence intervals.

 For the estimated retirement model, the distribution of durations of final nursing home stays is discretized into the seven intervals indicated in Appendix Table 3.10. For each age (by single years of age), health status, and condition group, the probabilities of the seven intervals is calculated according to the estimated distribution parameters in Appendix Table 3.9. The cost of being in each of these intervals is calculated as the average number of fitted months for those in the interval, given by the last row of Appendix Table 3.10, multiplied by the 1992 cost of that number of fitted months. For example, an individual dying with a particular age, health status, and condition group might be calculated to have a 20 percent chance of residing in a nursing home at the time of death and, conditional of residing in a nursing home, as 15 percent chance that his length of stay was 6 to 12 months. The cost of such an outcome would be calculated as 8.8 months times the average cost per month of being in a nursing home.

Appendix Table 3.1

Estimates of Social Security Disability Insurance Hazard

 Absolute

 Coefficient t-statistic

 Constant -2.504\*\*\* 16.53

 (Age – 60) / 10 0.045 0.27

 [(Age – 60) / 10]2 -0.424 0.83

 Health Category

 Good -0.532\*\*\* 3.61

 Poor 0.662\*\*\* 5.71

 Terrible 1.250\*\*\* 8.50

 Condition Group

 Group 2 0.398\*\* 2.22

 Group 3 0.484 1.42

 Group 4 0.069 0.40

 Group 5 0.406\*\*\* 2.63

 Group 6 0.346 1.58

 Group 7 0.314 1.45

 Group 8 1.542\*\*\* 3.31

 Number of observations: 7670

 Log-likelihood value: -429.05

Appendix Table 3.2

Percentage on Social Security Disability Insurance by Age

 Percent on SSDI

 Age Observed Simulated

 50 3.6 1.5

 51 0.0 1.7

 52 2.0 2.0

 53 1.4 2.3

 54 2.3 2.6

 55 3.7 3.0

 56 4.2 3.4

 57 4.5 3.9

 58 4.9 4.4

 59 4.4 4.9

 60 5.8 5.5

 61 4.6 6.0

 62 7.7 6.6

 63 6.3 7.1

 64 8.1 7.6

Appendix Table 3.3

Estimates of Log-Normal Distribution of Medical Costs

 Absolute

 Coefficient t-statistic

 μ 4.7402\*\*\* 70.88

 σ 1.7966\*\*\* 55.73

 Health Status

 Good -0.0717 1.10

 Poor 0.2011\*\*\* 3.09

 Terrible 0.3905\*\*\* 4.99

 Condition Group

 Group 1 -0.1876\*\*\* 2.68

 Group 2 0.3316\*\*\* 3.28

 Group 3 0.3769\*\* 2.27

 Group 5 0.3199\*\*\* 4.79

 Group 6 0.3784\*\*\* 3.56

 Group 7 0.3588\*\*\* 3.11

 Group 8 0.5107\*\* 2.00

 Number of observations: 5724

 Log-likelihood value: -19112.99

Appendix Table 3.4

Average Medical Care Expenditures by Percentile Range

 Percentile Range

 0-50 50-80 80-90 90-95 95-97.5 97.5-99 99-99.5 99.5-100

Health Status

 Good 37 230 695 1427 2632 4822 8509 23355

 Fair 44 272 821 1688 3112 5701 10059 27586

 Poor 59 367 1106 2272 4184 7658 13499 36936

 Terrible 76 470 1417 2908 5352 9787 17239 47090

Condition Group

 Group 1 34 209 627 1283 2356 4301 7563 20591

 Group 2 60 369 1108 2269 4170 7617 13403 36542

 Group 3 63 386 1162 2381 4377 7998 14079 38419

 Group 4 45 279 838 1719 3161 5777 10172 27769

 Group 5 66 411 1238 2539 4671 8539 15037 41055

 Group 6 68 423 1275 2615 4811 8798 15497 42342

 Group 7 72 446 1344 2755 5069 9267 16320 44563

 Group 8 86 535 1610 3301 6074 11103 19551 53374

Appendix Table 3.5

Percentage Who Reside in Nursing Homes at Time of Death

 Age Range

 50-59 60-69 70-79 80-89 90-99

Health

 Good 6.6 6.5 12.8 20.5 29.0

 Fair 2.2 8.0 14.4 27.2 50.9

 Poor 6.8 9.6 15.8 29.8 46.6

 Terrible 16.4 19.0 32.0 50.5 63.6

Condition Group

 Group 1 4.5 5.6 18.3 32.6 53.9

 Group 2 2.3 7.7 19.0 32.8 43.1

 Group 3 11.5 8.8 18.4 30.9 50.0

 Group 4 6.5 8.6 25.4 41.5 68.4

 Group 5 7.3 11.6 21.1 37.1 52.3

 Group 6 11.1 9.4 14.2 30.3 48.8

 Group 7 20.0 23.9 28.5 49.5 60.1

 Group 8 0.0 17.0 19.7 44.4 65.0

All 8.0 11.8 21.0 37.6 57.0

Appendix Table 3.6

Length of Nursing Home Stays in Months by Age Range

A. By Health Status

 Age Range

 50-59 60-69 70-79 80-89 90-99

 50th Percentile

Health Status: Good 3 0 1 2 12

 Fair 2 2 1 1 2

 Poor 3 1 2 3 4

 Terrible 10 3 10 12 18

All 3 2 3 5 10

 75th Percentile

Health Status: Good 4 1 3 5 36

 Fair 2 4 3 4 9

 Poor 3 3 7 10 12

 Terrible 24 21 24 24 36

All 12 10 12 22 27

 90th Percentile

Health Status: Good 4 3 5 13 36

 Fair 2 24 5 18 31

 Poor 108 8 12 24 36

 Terrible 60 48 52 48 68

All 60 36 36 36 52

Appendix Table 3 (cont.)

Length of Nursing Home Stays in Months by Age Range

A. By Condition Group

 Age Range

 50-59 60-69 70-79 80-89 90-99

 50th Percentile

Condition Group: 1 108 0 2 2 7

 2 0 1 2 4 5

 3 12 3 0 1 7

 4 3 1 3 4 12

 5 3 2 3 6 10

 6 2 1 2 4 9

 7 8 2 12 12 18

 8 a 24 6 6 6

 75th Percentile

Condition Group: 1 157 3 3 18 24

 2 0 3 4 22 16

 3 12 3 2 12 12

 4 24 1 17 24 36

 5 24 10 12 19 36

 6 3 2 6 16 24

 7 11 17 48 24 41

 8 a 108 12 24 36

 90th Percentile

Condition Group: 1 157 36 6 36 56

 2 0 3 14 48 24

 3 12 36 12 24 36

 4 24 12 36 60 60

 5 36 36 36 36 60

 6 3 8 12 36 48

 7 60 48 72 48 72

 8 a 144 60 37 60

 a No observations for this condition group in this age range

Appendix Table 3.7

Estimates of Nursing Home Probability Probit Equation

 Absolute

 Coefficient t-statistic

 Constant -0.956\*\*\* 11.98

 (Age - 75) / 10 0.405\*\*\* 23.10

 [(Age - 75) / 10]2 0.015 1.29

 Health Status

 Good -0.158\* 1.92

 Poor 0.063 1.16

 Terrible 0.530\*\*\* 9.87

 Condition Group

 Group 2 -0.092 0.91

 Group 3 -0.034 0.31

 Group 4 0.141\* 1.70

 Group 5 -0.085 1.11

 Group 6 -0.212\*\*\* 2.57

 Group 7 0.128 1.54

 Group 8 -0.007 0.07

 Number of observations: 7029

 Log-likelihood value: -3689.96

Appendix Table 3.8

Fitted Probabilities of Nursing Home Stays

 Age Range

 50-59 60-69 70-79 80-89 90-99

Health

 Good 3.4 6.4 12.9 22.8 34.4

 Fair 4.6 8.5 15.9 26.9 41.3

 Poor 5.2 9.3 17.2 29.7 45.0

 Terrible 12.6 20.3 33.0 48.9 63.7

Condition Group

 Group 1 4.8 8.2 17.9 31.2 50.3

 Group 2 4.5 8.7 17.0 31.3 46.9

 Group 3 6.1 9.0 18.7 32.3 50.3

 Group 4 7.9 14.1 24.4 43.2 61.2

 Group 5 6.7 11.9 21.2 35.9 53.1

 Group 6 4.6 9.3 17.1 30.2 47.0

 Group 7 11.8 19.8 31.9 48.0 63.3

 Group 8 8.2 16.5 27.7 41.4 57.9

All 6.6 12.3 21.4 36.0 53.3

Appendix Table 3.9

Estimated Parameters of Nursing Home Stay Distribution

 Absolute

 Coefficient t-statistic

 Probit for Stays of Less than Two Months

 Constant 0.2490 1.76

 (Age - 85) / 10 -2.0943\*\*\* 5.73

 [(Age - 85) / 10]2 -2.4443 1.10

 Health Status

 Good -0.1152 0.63

 Poor -0.1924\* 1.84

 Terrible -0.6884\*\*\* 6.92

 Condition Group

 Group 2 0.0203 0.11

 Group 3 0.1989 0.98

 Group 4 -0.0991 0.70

 Group 5 -0.1192 0.89

 Group 6 0.1065 0.72

 Group 7 -0.2969\*\* 2.08

 Group 8 -0.0924 0.54

 Exponential Distribution for Stays Longer than Two Months

 Constant 0.07320\*\*\* 10.38

 (Age - 85) / 10 -0.06483\*\*\* 5.35

 [(Age - 85) / 10]2 -0.27465\*\*\* 3.41

 Health Status

 Good 0.03175\* 1.75

 Poor -0.01749\*\*\* 2.82

 Terrible -0.04097\*\*\* 7.04

 Condition Group

 Group 2 0.01341 1.42

 Group 3 0.03260\* 1.65

 Group 4 0.00457 0.78

 Group 5 0.01046\* 1.84

 Group 6 0.01806\*\*\* 2.53

 Group 7 0.00192 0.34

 Group 8 0.00534 0.86

Appendix Table 3.10

Distributions of Lengths of Stays at Nursing Homes Prior to Death

 Length of Stay

 0-2 2-6 6-12 12-24 24-48 48-72 72+

Age Range Observed Distribution

 50-59 41.4 20.7 17.2 6.9 3.4 3.4 6.9

 60-69 57.4 13.5 8.1 7.4 8.1 1.4 4.1

 70-79 45.6 17.1 13.6 10.6 7.8 3.0 2.3

 80-89 38.6 17.0 11.4 17.4 10.4 3.3 1.9

 90-99 31.0 12.5 13.1 17.4 15.3 6.9 3.7

 Fitted Distribution

 50-59 55.4 7.9 8.9 11.1 9.9 3.8 3.0

 60-69 51.3 9.4 10.4 12.7 10.6 3.5 2.1

 70-79 49.7 10.3 11.3 13.4 10.5 3.2 1.7

 80-89 41.8 10.9 12.3 15.3 12.9 4.3 2.6

 90-99 33.8 10.4 12.3 16.4 15.7 6.2 5.2

 Fitted Average Length of Stay within Interval (Months)

 1.0 3.9 8.8 17.4 33.9 58.1 104.3

Appendix 4. Additional Simulations of the Retirement Model

 *Simulations Involving Health and Health Insurance.*

 The first results simply give a tabulation of simulations involving good health and poor health. Recall that there are around 10,000 simulations for each of the 2,231 individuals in the sample. We collect the results of all the simulations which yielded the top health status (good) at age 60 and all the simulations which yielded the next to worst health status (poor) at age 60 and present these results in Appendix Table 4.1. The survival statistics in this table are survival probabilities conditional on being alive at age 60. First, notice that those in good health at age 60 had an almost 4 year advantage in conditional life expectancy, 22.8 years vs. 19.0 years. Not surprisingly, those in good health at age 60 were more likely to be in one of the better health categories and less likely to be in one of the worse health categories at later ages than those in poor health at age 60. This effect would be more substantial if we considered that mortality strikes those in poor health more often than those in good health.

Appendix Table 4.1

Simulations of Health Status and Survival

Conditional on Health Status at Age 60

 Percent in Health Status Among Survivors

 Age ------------------------------------------------------------ Conditional

 Range Good Fair Poor Terrible Survivala

Good Health at Age 60

 60-64 76.9 18.8 3.8 0.5 96.6

 65-69 51.2 32.9 13.3 2.6 89.5

 70-74 39.9 35.5 19.6 5.1 77.3

 75-79 31.0 35.8 25.1 8.1 60.7

 80-84 23.6 34.7 30.0 11.8 41.4

 85-89 16.8 31.6 34.2 17.4 22.3

 90-94 11.7 27.4 36.5 24.4 7.5

 Life Expectancy at age 60: 22.8 years

Poor Health at Age 60

 60-64 6.2 20.8 63.8 9.2 89.1

 65-69 19.3 32.0 34.9 13.7 76.9

 70-74 20.7 32.8 32.7 13.7 61.7

 75-79 18.1 32.3 34.1 15.5 45.1

 80-84 14.2 29.7 36.1 19.9 28.6

 85-89 10.2 25.9 38.6 25.2 14.3

 90-94 7.3 21.9 38.7 32.1 4.4

 Life Expectancy at age 60: 19.0 years

 a Conditional on being alive at age 60

While Appendix Table 4.1 gives the differences between those in good health and poor health, it does not isolate the effect of current health status per se on future health status and mortality. This is because the simulations resulting in good health at age 60 are more likely to be associated with personal characteristics, such as not smoking or not being obese, which are associated with health transitions and mortality. To isolate the effect of current health status per se, we repeat the base simulation for the entire sample, giving everyone good health at age 60, and again giving everyone poor health at age 60. Each simulation has the same condition group at age 60 as in the base simulation, but afterwards the condition group transitions, health status, and mortality are governed by the estimated model. The results of these simulations are given in Table 4.2. These simulations suggest that the effects of health status per se tend to die out over time, with the difference between life expectancies differing only by about 1.4 years rather than the 3.8 years in the previous table. Much of the persistent differences in long-term effects of having good vs. poor health at age 60 thus appears to be due to other factors which are helping to determine the health status at age 60 in the first place and which continue to influence subsequent health transitions.

Appendix Table 4.2

Effect of Health Status at Age 60

On Subsequent Health Status and Survival

 Percent in Health Status Among Survivors

 Age ------------------------------------------------------------ Conditional

 Range Good Fair Poor Terrible Survivala

Changing Health Status at Age 60 to Good

 60-64 73.1 21.2 5.0 0.7 95.7

 65-69 44.6 34.7 16.7 3.9 87.7

 70-74 34.0 35.7 23.2 7.1 74.3

 75-79 26.5 35.1 28.1 10.3 56.9

 80-84 20.2 33.2 32.2 14.4 37.8

 85-89 14.5 29.9 35.7 19.9 19.8

 90-94 10.2 25.7 37.2 26.8 6.4

 Life Expectancy at age 60: 21.9 years

Changing Health Status at Age 60 to Poor

 60-64 9.5 23.9 59.7 6.9 91.3

 65-69 28.8 33.2 28.4 9.7 81.2

 70-74 29.9 34.1 26.5 9.5 67.7

 75-79 25.3 34.4 29.1 11.3 51.6

 80-84 19.8 32.9 32.5 14.8 34.2

 85-89 14.3 29.7 35.8 20.2 18.0

 90-94 10.1 25.6 37.3 27.0 5.8

 Life Expectancy at age 60: 20.5 years

 a Conditional on being alive at age 60

In the following table we use the model to address the effects of an elimination of diabetes. The first step is to trace through the effects of eliminating diabetes in the health sub-model. These results are presented in Appendix Table 4.3.

Appendix Table 4.3

Effects of Diabetes on Simulated Percentages in Various Health States by Age Range

 Percent in Health Status Among Survivors

 Age ------------------------------------------------------------ Conditional

 Range Good Fair Poor Terrible Survivala

 Base Case

 50-54 59.1 25.2 10.8 5.0 96.3

 55-59 52.2 28.7 14.4 4.6 92.3

 60-64 46.2 31.4 17.2 5.2 86.6

 65-69 39.8 33.5 20.4 6.3 78.5

 70-74 33.1 34.9 24.1 7.9 66.1

 75-79 26.4 34.8 28.2 10.5 50.6

 80-84 20.3 33.2 32.2 14.4 33.7

 85-89 14.6 29.9 35.6 19.8 17.7

 90-94 10.3 25.8 37.2 26.8 5.7

 Life Expectancy at age 60: 21.4 years

 With Diabetes Eliminated

 50-54 59.4 25.0 10.7 5.0 96.4

 55-59 52.6 28.5 14.2 4.6 92.6

 60-64 46.7 31.2 16.9 5.2 87.1

 65-69 40.4 33.2 20.1 6.3 79.1

 70-74 33.7 34.7 23.7 7.9 66.9

 75-79 27.0 34.6 27.9 10.4 51.4

 80-84 20.9 33.0 31.8 14.2 34.4

 85-89 15.2 29.8 35.3 19.7 18.2

 90-94 10.8 25.7 36.9 26.6 6.0

 Life Expectancy at age 60: 21.5 years

 a Conditional on being alive at age 50

The effects of an elimination of diabetes on work effort are presented in Appendix Table 4.4. Given the lack of much of an effect on health transitions, the elimination of diabetes has only a minor effect on work effort, increasing the average retirement age by only a week or so.

Appendix Table 4.4

Effects of Diabetes on Work Effort

 Simulation Percentage

 With Point Percent

 Diabetes Base Increase Increase

 Age Eliminated Simulation in Work in Work

 Percent Working Full-Time

 54 79.8% 79.7% 0.1% 0.1%

 55 75.9 75.9 0.0 0.0

 56 72.8 72.7 0.1 0.1

 57 68.9 68.8 0.1 0.1

 58 64.7 64.5 0.2 0.3

 59 60.4 60.2 0.2 0.3

 60 54.8 54.6 0.2 0.4

 61 50.5 50.3 0.2 0.4

 62 36.4 36.2 0.2 0.6

 63 32.2 32.0 0.2 0.6

 64 27.0 26.8 0.2 0.7

 65 21.0 20.8 0.2 1.0

 66 17.5 17.3 0.2 1.2

 67 14.0 13.8 0.2 1.4

 Percent Working At All

 54 84.9 84.8 0.1 0.1

 55 81.7 81.6 0.1 0.1

 56 79.1 79.0 0.1 0.1

 57 76.1 76.0 0.1 0.1

 58 72.3 72.1 0.2 0.3

 59 68.3 68.1 0.2 0.3

 60 63.5 63.3 0.2 0.3

 61 59.7 59.5 0.2 0.3

 62 51.2 50.9 0.3 0.6

 63 47.7 47.4 0.3 0.6

 64 42.9 42.6 0.3 0.7

 65 37.2 36.9 0.3 0.8

 66 33.8 33.5 0.3 0.9

 67 30.1 29.9 0.2 0.7

*Simulations for Different Lifetime Income Groups.*

 Appendix Table 4.5 gives the evolution of health status and survival for each of these income groups. The higher income group has a life expectancy over two years longer than the lower income group, and they are more likely to be in one of the better health categories and less likely to be in one of the worse health categories at every age. It should be emphasized that these results are not a direct effect of income on health status and survival, since the transition equations for health status and survival do not include income as an explanatory variable. Rather, these differences arise from the effects of differing personal characteristics such as smoking, obesity, and so forth among the income groups.

Appendix Table 4.5

Simulations of Health Status and Survival by Income Group

 Percent in Health Status Among Survivors

 Age ------------------------------------------------------------ Conditional

 Range Good Fair Poor Terrible Survivala

Low Income

 50-54 55.6% 26.0% 12.3% 6.1% 95.7%

 55-59 48.4 29.5 16.2 5.8 91.2

 60-64 42.5 31.9 19.1 6.5 84.8

 65-69 36.4 33.5 22.3 7.8 75.7

 70-74 30.1 34.5 25.8 9.6 62.4

 75-79 23.9 34.0 29.8 12.3 46.2

 80-84 18.3 31.9 33.4 16.4 29.3

 85-89 13.1 28.4 36.3 22.2 14.1

 90-94 9.2 24.2 37.2 29.5 3.7

 Life Expectancy at age 60: 20.3 years

Middle Income

 50-54 58.3 25.6 11.0 5.1 96.2

 55-59 51.3 29.2 14.8 4.8 92.1

 60-64 45.2 31.8 17.7 5.4 86.4

 65-69 38.8 33.7 20.9 6.5 78.1

 70-74 32.1 34.9 24.7 8.3 65.7

 75-79 25.5 34.5 28.9 11.0 50.0

 80-84 19.6 32.7 32.7 15.0 32.9

 85-89 14.0 29.2 36.0 20.7 16.9

 90-94 9.9 25.1 37.2 27.9 5.2

 Life Expectancy at age 60: 21.2 years

High Income

 50-54 62.7 24.2 9.3 3.8 96.7

 55-59 56.2 27.7 12.6 3.5 93.2

 60-64 50.0 30.8 15.2 4.0 88.3

 65-69 43.4 33.4 18.3 4.9 81.0

 70-74 36.2 35.4 22.1 6.4 69.7

 75-79 29.0 35.7 26.5 8.8 54.9

 80-84 22.3 34.4 30.9 12.4 38.1

 85-89 16.0 31.4 34.9 17.7 21.4

 90-94 11.1 27.1 37.2 24.6 7.9

 Life Expectancy at age 60: 22.4 years

 a Conditional on being alive at age 50

Appendix Table 4.6 shows the simulated full-time work effort by age for the three lifetime income groups. There is a slight tendency for the low income group to work full-time somewhat more before the age of 62, and somewhat less after that age. These two differences roughly offset one another, however, and the average simulated retirement age for all three groups is very nearly equal.

Appendix Table 4.6

Simulations of Full-Time Work Effort for the Income Groups

 Income Group

 ---------------------------------------------------

 Age Low Medium High

 54 78.6% 80.7% 79.8%

 55 76.6 75.5 75.6

 56 73.4 73.0 71.8

 57 70.6 68.4 67.6

 58 67.0 63.7 63.2

 59 64.1 59.3 57.8

 60 59.5 53.7 51.3

 61 56.7 49.1 45.9

 62 31.8 37.3 39.0

 63 27.4 33.7 34.5

 64 22.9 28.2 28.8

 65 19.2 21.2 21.7

 66 16.0 17.7 18.0

 67 12.6 14.3 14.4

 Appendix Table 4.7 reports on the effects of providing universal insurance to the group of individuals who are uninsured. This group includes both individuals who are uninsured in their main full-time jobs and individuals who are insured in their main full-time jobs but uninsured after they left those jobs. The difference between Appendix Table 4.7 and Table 11 is that Appendix Table 4.7 includes the group who are uninsured in their main full-time jobs. Job lock is not an issue for these individuals, and the job lock incentives to remain in their main jobs to remain insured are not present. As a consequence, adding health insurance both before and after retirement for these individuals has much less of an employment effect than it does for individuals who lose health insurance upon retirement. The numbers in the first four columns of Appendix Table 4.7 are generally somewhat lower than the comparable numbers in Table 11, reflecting that this group is a mix of some individuals who are subject to job lock and other individuals who are not.

 The last two columns of Appendix Table 4.7 compare the effects of universal health insurance on retirement for two groups, those in good health and those in poor health. One might expect that the incentives for employees subject to job lock to remain working would be greater for those in poor health, since after retirement those in poor health are more likely to face substantial medical expenses. The simulated results suggest that the employment responses of those in poor health are indeed larger, especially at younger ages, but the differences between the results for those in good health and those in poor health are modest at best.

Appendix Table 4.7

Effects of Insuring the Uninsured

For Uninsured Sample

 By Income Group By Health Status

 Uninsured ------------------------------------------ ------------------------

Age Sample Low Medium High Good Poor

Percentage Point Change in Retirement from Full-Time Work

54 0.0 0.2 0.1 0.0 0.1 0.1

55 0.0 0.1 0.0 0.1 0.0 0.1

56 0.1 0.0 0.0 0.1 0.0 0.3

57 0.1 0.1 0.0 0.1 0.0 0.2

58 0.1 0.0 0.1 0.2 0.1 0.2

59 0.0 -0.2 0.2 0.3 0.0 0.2

60 0.2 -0.2 0.3 0.3 0.2 0.2

61 0.2 -0.1 0.4 0.4 0.2 0.1

62 0.3 0.1 0.6 0.2 0.3 0.3

63 0.3 0.2 0.5 0.2 0.3 0.3

64 0.3 0.3 0.4 0.2 0.3 0.3

65 0.2 0.2 0.3 0.1 0.2 0.2

66 0.0 0.1 0.1 0.0 0.0 0.1

67 0.0 0.1 0.1 0.0 0.0 0.1

Percentage Point Change in Full Retirement

54 0.0 0.1 0.0 -0.1 0.1 0.0

55 0.0 -0.1 0.0 0.0 -0.1 0.0

56 0.0 0.0 -0.1 0.1 -0.1 0.1

57 0.0 0.0 -0.1 0.1 0.0 0.0

58 -0.1 -0.2 0.0 0.1 -0.1 0.0

59 -0.1 -0.3 0.0 0.2 -0.1 0.0

60 0.0 -0.3 0.3 0.1 0.0 0.0

61 0.1 -0.1 0.3 0.1 0.1 -0.1

62 -0.1 -0.3 0.2 0.0 -0.1 0.0

63 0.1 -0.2 0.3 0.2 0.1 0.0

64 0.1 0.0 0.3 0.2 0.1 0.1

65 0.3 0.3 0.3 0.2 0.2 0.2

66 0.0 0.0 0.0 -0.1 0.0 0.0

67 -0.1 0.0 0.0 0.0 0.0 0.0

Appendix Table 4.8 shows the results of providing universal insurance over the entire sample, including those who had retiree coverage. In the model, adding universal insurance for those who already had retiree coverage does not change their situation and has no effect, so adding this group to the sample simply dilutes the results relative to Appendix Table 4.7. Again, remember that the sample is married males with substantial work histories in the original HRS cohort, and most of those individuals had some form of insurance after they retired, either from their previous work or from some other source.

Appendix Table 4.8

Effects of Insuring the Uninsured

For Entire Sample

 By Income Group By Health Status

 Entire ------------------------------------------ ------------------------

Age Sample Low Medium High Good Poor

Percentage Point Change in Retirement from Full-Time Work

54 0.0 0.0 0.1 0.0 0.0 0.0

55 0.1 0.1 0.1 0.0 0.0 0.0

56 0.0 0.0 0.0 0.0 0.0 0.1

57 0.0 0.0 0.0 0.0 0.0 0.1

58 0.0 0.0 0.1 0.1 0.0 0.1

59 0.0 0.0 0.0 0.1 0.0 0.0

60 0.0 0.0 0.1 0.0 0.0 0.1

61 0.0 -0.1 0.1 0.0 0.1 0.0

62 0.1 0.0 0.1 0.0 0.0 0.1

63 0.0 0.0 0.1 0.0 0.0 0.1

64 0.1 0.1 0.1 0.0 0.1 0.0

65 0.1 0.1 0.0 0.0 0.1 0.0

66 0.0 0.0 0.0 0.0 0.0 0.0

67 0.0 0.1 0.1 0.0 0.0 0.0

Percentage Point Change in Full Retirement

54 0.0 0.0 0.0 0.0 0.0 0.0

55 0.0 0.0 0.0 0.0 0.0 0.0

56 0.0 0.0 0.0 0.0 0.0 0.0

57 0.0 0.0 0.0 0.0 0.0 0.0

58 0.0 -0.1 0.0 0.0 0.0 0.0

59 0.0 -0.1 0.0 0.0 0.0 0.0

60 0.0 -0.1 0.0 0.0 0.0 0.0

61 0.0 0.0 0.1 0.1 0.0 0.0

62 0.0 -0.1 0.0 0.1 -0.1 0.1

63 0.0 0.0 0.1 0.1 0.1 0.0

64 0.0 0.0 0.0 0.1 0.1 0.0

65 0.0 0.1 0.0 0.0 0.0 0.1

66 0.0 0.0 0.0 0.0 0.0 0.0

67 0.0 0.0 0.0 0.0 0.0 0.0

1. There are no entries in the -1.4 to -1.0 range. The entries in the -1.5 to -1.4 range are the observations where all five of the health sub-measures are perfect. If any of them are not perfect, the equations push the health measure above the -1.0 mark. [↑](#footnote-ref-1)