NBER WORKING PAPER SERIES

EVALUATION OF SUBJECTIVE PROBABILITY DISTRIBUTIONS IN THE HRS

Michael Hurd
Kathleen McGarry

Working Paper No. 4560

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December, 1993

Financial support from the University of Michigan and the National Institute on Aging is gratefully acknowledged. Many thanks to Nancy Cole for research assistance. This paper is part of NBER's research program in Aging. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Working Paper #4560
December 1993

EVALUATION OF SUBJECTIVE PROBABILITY DISTRIBUTIONS IN THE HRS

ABSTRACT

In the Health and Retirement Survey respondents were asked about the chances they would live to 75 or to 85, and the chances they would work after age 62 or 65. We analyze the responses to determine if they behave like probabilities, if their averages are close to average probabilities in the population, and if they have correlations with other variables that are similar to correlations with actual outcomes. We find that generally they do behave like probabilities and they do aggregate. Most remarkable, however, is that they covary with other variables in the same way actual outcomes vary with the variables. For example, smokers give lower probabilities of living to 75 than nonsmokers. We conclude that these measures of subjective probabilities have great potential use in models of intertemporal decision making under uncertainty.

Michael Hurd
Department of Economics
SUNY, Stony Brook
Stony Brook, NY 11794
and NBER

Kathleen McGarry
Department of Economics
UCLA
Los Angeles, CA 90024-1477
and NBER

1. Introduction

Many economic models are based on forward looking behavior by economic agents. Although it is often said that "expectations" about future events are important in these models, more precisely it is the probability distributions of future events that enter the models. For example, consumption and savings decisions of an individual are thought to depend on beliefs about future interest rates, the likelihood of dying, and the risk of substantial future medical expenditures. According to this theory, decision makers have subjective probability distributions about these and other events and they use them to make decisions about saving.

In a few microeconomic models, we have data on probability distributions that may plausibly be assumed to approximate those required by the models of decision making under uncertainty. For example, life cycle models of consumption in which mortality risk helps determine saving have been estimated by assuming that individuals have subjective probability distributions on mortality risk that are the same as those found from life tables. In most applications, however, we do not have data on probability distributions, so estimation requires some unverifiable assumptions. For example, in macroeconomic models expectations are assumed to be rational, which often yields an estimable relationship; yet, the rationality assumption cannot be tested outside of the context of the model. In life cycle models of saving, the average mortality risk of a cohort may not be well approximated by the life table mortality risk because of changing risk: a cohort may not believe that the mortality experience of older cohorts is the same as its will be. Furthermore, individuals within a cohort will have different subjective probability distributions on mortality risk because of observable and unobservable differences in mortality risk factors. Finally, an individual's own subjective evaluation of probability distributions determines behavior, even if it is systematically incorrect; yet that evaluation is not generally observable.

The Health and Retirement Survey (HRS) has a number of innovative questions in which respondents were asked on a 0-10 scale the chances of future events such as working full-time past age 62 or living to age 75.[1] After rescaling to 0-1, these can be interpreted as subjective probability distributions on the events. They have the potential to change substantially the way in which we estimate stochastic dynamic models based on micro data because they can supply probabilities of events for which we have no population averages, and because they contain individual heterogeneity about probabilities. They can, in principle, be used directly in our models of decision making. This makes them different from subjective evaluations that have been elicited in previous surveys: questions such as "When do you expect to retire?" have been asked before, but the responses cannot be used in a quantifiable way in our models.

While the HRS questions about subjective probabilities have great potential, it is certainly possible that, as an empirical matter, they are not particularly useful. For example, respondents may have little idea of the probabilities of future events, or they may answer at random. Of course the best evaluation of them will come from a comparison of the probabilities with outcomes in the panel. But even in cross-section we can learn a great deal.

The broad goal of this paper is to evaluate the subjective probability distributions in three ways. First, we will check external consistency: how do the probabilities compare with probabilities found in external data? We compare averages of the subjective probability distributions with population averages such as probabilities of living from life tables and retirement probabilities. Second, we study the internal consistency of the subjective probability distributions to see if they behave like probabilities. For example, do they imply conditional probabilities that are between zero and one.

Our third kind of evaluation takes the subjective probabilities to be actual outcomes. We explain those outcomes with simple equations estimated over individual level data and compare the estimates with results from the literature.

[1]See Juster and Suzman (1993) for a description of the HRS.

2. Measures of subjective probability distributions in the HRS.

The HRS has a number of questions that can be interpreted as subjective probability distributions. All the questions are asked in the following form:

> "Using any number from zero to ten where 0 equals absolutely no chance and 10 equals absolutely certain, what do you think are the chances you will live to be 75 or more?"

00	01	02	03	04	05	06	07	08	09	10

Absolutely no chance — Absolutely certain

> "85 or more?"
>
> "You will be working full-time after you reach age 62? Age 65?"

and other questions on housing purchase, job stability, financial help to family, housing prices, Social Security, and the economy.

In this paper we study the responses to the questions about living to 75 or 85, which we will call Plive75 and Plive85, and the responses the questions about working full-time, which we will call Pwork62 and Pwork65. After normalizing to [0,1] we will call these the probabilities of living or of working, but they are, at best, measures of subjective probabilities. We have chosen to focus on these probabilities because much more is know about what constitutes reasonable answers than to the other subjective probabilities both with respect to level and to how they covary with other observable data.

3. Probabilities of living to 75 or 85.

For population comparisons, our sample is restricted to the age range 51-61, and we use sampling weights to account for oversampling of blacks, Hispanics and Floridians. For analysis we use a sample of men aged 51-65 and women aged 46-61, and who were not represented by a proxy interview. We realize that outside of the age range 51-61, the sample is not representative of the population because a respondent must be a spouse of an age-eligible person. Nonetheless, we wanted more age variation than in the age-eligible sample, particularly because we want to find how the subjective probabilities vary as age approaches 65 or 75. Furthermore, about 23% of the sample is outside the age range 51-61, which is a large fraction to drop in the absence of a compelling reason.[2]

We have 7946 observations that we will use for the results in this section. This is based on the responses to Plive75. (We have slightly fewer responses to Plive85). The response rate in the entire survey to Plive75 and Plive85 is about 98%.

3.1. Comparisons with life tables

We begin by comparing in Table 1 Plive75 and Plive85 with averages from life tables. Plive85 is less than Plive75, and Plive85 given Plive75, is 0.66. The levels of Plive75 averaged over men and women are close to the averages in the 1988 life table, but the Plive85 are higher than those from the life table[3] Taking the life table as the relevant comparison, men substantially over-estimate the probability they will live to 85, and women under-estimate the probability they will live to 75. As a consequence, both over-estimate the conditional probability of living to 85 given alive at 75.

There have been substantial reductions in mortality risk over a number of years, and the reductions are expected to persist. It is relevant, therefore, to wonder how people form their expectations about the length of life and how the expectations might vary from cohort to cohort. The second part of the table has estimates of the probability

[2]For analysis we often would like to know if a model seems to hold for any population provided the population was not chosen either to favor or disfavor the model. Based on this reasoning we imagine that most analyses will be able to use the part of the HRS outside of the age range 51-61.

[3]The life table averages are weighted age-specific estimated probabilities, where the weights are the number of men or women at each age in our HRS sample. The age range is 51-61.

of living to 75 or 85 from age 55. The last three lines come from, respectively, a 1980 life table (based on observed age-specific mortality rates in 1980), a 1988 life table, and a 2000 life table, which is, of course, based on forecasts of changes in mortality risk. The changes are substantial, which makes it difficult to know what is a good standard of comparison: the 1988 life table is the product of age-specific mortality rates in 1988, which could be quite different from the age-specific mortality risks the HRS population anticipates. From this point of view, even the "overestimate" of Plive85 by men could be a reasonable projection. At a minimum we would expect the HRS sample to give higher rates of Plive75 and especially Plive85 than the 1988 life table because of cohort effects.

Figure 1 has the distributions of Plive75 and Plive85. They have considerable bunching at 0, 0.5 and 1.0. An interpretation is that people choose one of the three points according to whether they are rather confident, not confident at all, or uncertain about living to 75 or 85. However, there are mini-spikes at 0.2 and 0.8, and particularly for Plive85, considerable mass at other points. In our view the distribution cannot be reduce much further without the possible loss of considerable information.

In Figure 2, we have, for the moment, extended our sample to include men aged 46 to 74, and, in Figure 3, women aged 38-65.[4] We did this to get the greatest possible age range. As a reminder of the thin sample at ages far from the HRS age range, we show the distribution of observations at the bottom of the graph. The averages by age of Plive75 and Plive85 are compared with estimates of Plive75 and Plive85 from the 1988 life table. As we saw earlier Plive85 is considerably greater than the life table estimates. What is most notable is that the age-paths of Plive75 and Plive85 are rather flat except at ages above about 64, when they rise rather sharply. Figure 3 shows age paths of Plive75 and Plive85 for women. The paths are flat and possibly even declining before 50.

If mortality risk is stationary over time and there were no heterogeneity in the population, these paths should slope upward, reflecting that the probability of dying in any year is positive. It is unlikely, however, that either of these conditions is met.

[4]We only use the extended sample for these two graphs.

Figure 4 shows Plive75 as a function of age estimated from the 1980, 1988 and 2000 life tables, and Figure 5 shows Plive85. It is certainly reasonable to imagine that the upward shift has been so rapid that cohorts in their early 50's could have conditional probabilities of living to 75 that are about the same as older cohorts. For example, if a 50 year-old man uses the year 2000 life table to form probabilities, his estimate of Plive75 would be about 0.6. If a 62 year-old uses the 1988 life table his estimate would be about 0.62.

At greater ages the slope of the conditional probability curve becomes steep, so that the shift is not important, and the cohort effect will be negligible. The age pattern of Plive75 implied by this example is about what we saw in Figure 2. The effect of the upward shift is even greater on Plive85 (Figure 5). The same comparison we just made would lead to a declining Plive85 in cross-section data. This is roughly what we observed in Figures 2 and 3.

We do not know how people form their subjective probabilities about living to 75 or 85. But the rapid change in mortality risk leads us to conclude that a declining path of Plive75 and Plive85 with age, especially at younger ages, can be consistent with our thinking of them as probabilities.

3.2. Internal consistency

Although we may not want to make predictions about how Plive75 or Plive85 vary across individuals by age, we can make predictions about the relationship between them at the individual level: in that each individual has a positive probability of dying between 75 and 85 should he live to 75, Plive75 should be greater than Plive85 for each individual. Figure 6 has estimates of the mean of Plive85 given Plive75. It is just the average of Plive85 over all those who gave a particular value of Plive75. For reference the figure shows the 45 degree line. The graph shows that Plive85 given Plive75 is less than Plive75, and the difference increases with Plive75. Therefore, on average Plive85 given Plive75 lies between zero and one.

Table 2 has information about the joint responses. These figures and those in the rest of this section are based on our sample of women 46-51 and men 51-65. About 70% of the individuals have Plive75 greater than Plive85. If we add in the zeros as reasonable answers, we have about 77% of the sample whose responses satisfy either Plive75 > Plive85 or both probabilities are zero. It is not clear how much one should be disturbed by the other cases: the ties could be explained by uninformed guessing by the respondent or observation error, which would have to be modelled by an analyst. We find the 2.5% with Plive75 < Plive85 and the 9.2% with Plive75 = Plive85 = 1.0 more worrisome: they constitute 11.7% of the sample that may not have understood the nature of the question. However, the response rate to the probability questions is very high, and it should be clear that there is information even in these lowest quality responses. All variables derived from household interviews have inconsistencies and observations error. We conclude that the inconsistencies in Plive75 and Plive85 are tolerable and that their inconsistencies and errors are probably no larger than those of many other variables such as household wealth.

Figures 7-10 have examples of the conditional distribution of Plive85. Figure 7 gives the distribution of Plive85 given that Plive75 = 0.2. In about 90% of the cases Plive85 is less than 0.2. Figure 8 (Plive75=0.5) has some bunching at 0.5: apparently in the face of considerable uncertainty some simply answered 50-50 to the questions. Again, this does not mean the responses have no information.

As shown in Figure 9, most of the respondents who gave Plive75=0.8 gave smaller values for Plive85; just 2.8% gave higher values.

Fig 10, which has the responses over those with Plive75=1.0 (22% of the sample), shows that a large percentage of them, 42%, said Plive85 is also equal to 1.0. It is, of course, possible that these are optimistic people, and that they act as if their probabilities of surviving to 75 or 85 are close to one. However, it is quite likely that some did not understand the question, and, had they not been bound by the scale (which they had as a visual cue) they would have answered with numbers larger than one. We shall have to wait in the panel to see if these probabilities change with changes in life events.

We thought that with age the responses might become more heterogeneous: as people get new information about their health status and as they age toward 75, they may either become convinced they will live to 75 or convinced they will not live. Thus, rather than the average being composed of everyone having the population probability, it would be composed of a fraction with probability one and another fraction with probability zero so that the fractions averaged to the population probability. If such a process happens with age the variance of Plive75 and Plive85 should increase with age. We studied the variation in the standard deviation of Plive75 and Plive85 as a function of age, but we could not see any pattern.[5]

We conclude that, although there is some internal inconsistency, broadly speaking the observations on Plive75 and Plive85 act like probabilities, and, given the changes in life tables over time, they aggregate to reasonable levels.

3.3. Covariation with other variables.

Even with changing mortality risk from cohort to cohort, at least the sign of the variation of Plive75 and Plive85 with risk factors should remain constant. For example, someone who smokes should have a lower probability of living to 75 than someone from the same cohort who does not smoke. Averaging the probabilities over smokers and nonsmokers will reveal that difference.[6] A difference will be found after averaging over cohorts unless the incidence of smoking varies substantially with cohorts and there is a change in cohort-specific mortality risk. It is beyond the scope of this paper to study changes in risk factors by cohort, so we will assume that the incidence of risk factors is roughly constant.

In the analysis to follow we will find differences in Plive75 and in Plive85 as risk factors vary. It will help our understanding if we can translate a change in a probability into something with which we are more familiar such as a change in life expectancy. For

[5]This is complicated by the mean and variance of the probabilities not in general being independent: just as in binomial sampling, the maximum variance is at a probability of 0.5; but, of course, the variance in Plive75 and Plive85 will be zero if everyone has a probability of 0.5.

[6]A difference in probabilities does not require a causal relationship between smoking and longevity, only that the correlations among observed and unobserved variables be similar in the population and in the HRS.

example, in the 1988 life table Plive75 is 0.59 for the men in our sample. Suppose it were 0.03 higher among nonsmokers than among nonsmokers. Is this difference large? We can get a rough idea in terms of live expectancy as follows. Among men, the probability of living to 75 from 55 is about 0.59 and of living to 74 is 0.62. If a 55 year-old man has a Plive75 of 0.62, he believes his chances of living to 75 are the same as the chances of a random 55 year-old living to 74. A way to construct his individual life table is simply to shift the population life table by a year so that the population probability of living to 74 is now his probability of living to 75, and to make the probability that he will live to 56 equal to 1.0. Then his life expectancy, the integral under the life table function, is the integral from age 56 up plus the integral from 55 to 56. The first part is the same as the life expectancy of the 55 year-old population and the second part is 1.0. Therefore, the change in life expectancy is one year. We will take this approximation.

In that the life expectancy of a 55 year-old male is about 22 years, a 0.03 change in Plive75 on a base of 0.59 will change life expectancy by about 5%. Among 55 year-old women, a change of 0.02 in Plive75 (on a base of 0.74) will change life expectancy by a year, which is about 4%. Both of these "elasticities" are fairly close to one. Although a change in Plive85 of 0.03 is a larger proportionate change than in Plive75, its effect on the life expectancy of men is about the same, one year. Among women, because the Plive85 curve has a steeper slope, an increase of a year in life expectancy requires an increase of about 0.04 in Plive85. We will use these rough approximations as a guide in assessing the importance of a change in Plive75 or Plive85 as risk factors change.

It is well known that mortality risk varies with a number of indicators of socio-economic status: education, wealth and income to name but a few. Table 3 has Plive75 and Plive85 by wealth quartiles, and indeed, the variation is substantial: taking a linear extrapolation of the relationship between Plive75 and life expectancy the difference in life expectancies between the first and fourth quartiles is abut five years; according to Plive85 it is about two years.

The variation by education level is about the same (Table 4) with approximately the same implications for life expectancy.

The HRS respondents were asked to give a self-assessment of their health. Table 5 has the distribution of responses in our working sample and the averages of Plive75 and Plive85 by health status. The variation is enormous: Plive75 ranges from 0.34 to 0.75 among men and 0.40 to 0.78 among women, and with similar variation in Plive85. This is roughly a difference in life expectancy of 13 years at age 55.

Within health categories, Plive75 and Plive85 are higher among women than among men. Women have fewer risk factors such as smoking, and there are surely other unobserved determinents of longevity that vary by sex even holding constant health status.

Tables 6 and 7 show the life probabilities as a function of smoking and drinking. Smoking, of course, is a risk factor in the population, and that is found in Plive75 and Plive85. Furthermore, the difference between "never smoked" and "not now" (but in the past) is rather small just as it is in the population. In epidemiological data, moderate drinking is associated with greater longevity, and heavy drinking (five or more drinks per day) with substantially lessened longevity. This is precisely what is found in Plive75 and Plive85.

Table 8 has the averages of the life probabilities by health and education. Because of the positive correlation between health and education the effects of education on the life probabilities is much smaller than when health is not kept constant. For example, at health levels of good or very good (where most of the observations lie) Plive75 varies by just 0.03 or 0.04 with eduction level, and Plive85 be even less. Yet, within education level they vary with health status by about as much as they do in Table 5.

We have similar results when we interact health status with smoking status, or with income or wealth quartiles. Within health categories the variation in the probabilities of living is much smaller than in Tables 3 and 6; yet, among smokers or within an income or wealth quartile, the probabilities vary substantially with health status. Apparently the main result of smoking is to change self-assessed health (and actual health), which, in turn, changes life expectancy. The main effect of income or wealth is to signal differences in health status.

Within the health categories "good" or "very good" moderate drinking has only very small effects (Table 9). Yet, the difference between heavy drinking and moderate drinking is about as large in Table 7, which has no control for health status. Possibly heavy drinkers whose health is very good anticipate a decline in health status and they have incorporated that into their subjective probability distributions. The overall variation in Plive75 and Plive85 in the table is remarkably large, and the variation is both internally and externally consistent. For example, Plive75 is 0.74 among those who do not drink at all and are in excellent health; it is 0.30 among those who are heavy drinkers and in poor health. This kind of variation (large and consistent) increases our confidence that these subjective probabilities will be good predictors of actual mortality outcomes.

Although the cross-tabulations are suggestive and in accord with what is known from epidemiological data, we would like to know better the source of variation in Plive75 and Plive85. We do this with linear regressions of Plive75 and Plive85 on measures of socio-economic status, personal characteristics, risk factors, diseases, and self-assessed health status.

Table 10 has two sets of estimated coefficients. The first has observable variables on the right-hand side; the second has, in addition, health status. We first discuss the coefficients in the first column, the results when health status is excluded.[7]

Income has a small, not significant coefficient; wealth has a small coefficient with a t-statistic just over 1.95. We say these are small in that the variation in Plive75 explained by the coefficients on income and wealth as income and wealth vary across quartiles is small compared with the actual variation across the quartiles (Table 3).

The change with age is much smaller than what is found in a life table: here 10 years change Plive75 by about 0.04 compared with about 0.15 in the 1988 life table. We have already discussed how cohort effects could account for the difference.

The measures of physical activity (normalized at three or more per week)

---

[7]The standard errors do not change by much when health status is excluded, so we do not report them in the table.

apparently classify people into those who are physically activity and those who are not. Not being physically active reduces Plive75 by 0.04 to 0.06, and it matters little if the physical activity is light or heavy. The result, of course, does not imply that exercise will increase longevity because health status will influence both whether people are physically active and longevity.

Both smoking, drinking and education have smaller effects than in the cross-tabulations.

The incidence or prevalence of diseases affects Plive75 as would be expected and the effects are large: all are negative, and many of them reduce life expectancy by 2-4 years in our metric. For example, ever having had cancer or malignant tumor reduces Plive75 by 0.072, which reduces life expectancy for women by about 3.6 years. To the extent smoking affects life expectancy by causing these diseases, including them in the regression will attenuate the effects of smoking, which is what we observe.

Adding the health variables increases the R^2 from 0.094 to 0.158: apparently people use information that is not observable in answering both the question about health status and about the subjective probability distributions.

Income and wealth now have very small coefficients (second column of Table 10). The physical activity variables are not very important with the exception of the difference between never having any heavy physical activity and having some more than once a month. As before, we imagine that the difference is not causal, but simply reflects that people who are not physically active often are not able to be physically active. Never having physical activity simply provides finer detail on a measure of health than the five categories "excellent" to "poor."

The health variables have very large coefficients: Plive75 differs by 0.35 between excellent and poor health even though we have a number of socio-economic variables and nine disease indicators in the regression. This is probably about half of life expectancy at age 55.

Among the diseases all the coefficients are smaller in absolute magnitude, and only cancer has a significant coefficient. Apparently most of the effect of diseases on the subjective probabilities works through their effect on self-assessed health.

Table 11 has corresponding results for Plive85. The effects of the variables are generally smaller than in the estimated regression of Plive75. The exception is diseases: for example the effect of ever having had heart problems is -.069 on Plive75 and -.097 on Plive85. As before, including the health variables reduces the importance of the other risk factors.

Genetic factors also help determine life expectancy, and the age at which parents die is an important indicator of the genetic predisposition to longevity. We imagine, however, that the functional relationship between the parents' age at death and the child's mortality risk is rather complicated. In that the leading cause of death at an early age is accidents, the effects of the very early death of a parent on Plive75 or Plive85 will probably be qualitatively different from the effects of a later death. In particular, the effect will not be monotonic in the age of the parents' death. We allow for this with a set of categorical and continuous variables in each parent's age, if alive, and in each parent's age at death, if dead.

Table 12 has selected coefficients from a regression that includes parents' age if alive and parents' age at death if dead. The table also shows the distribution of the parents' mortality status.[8]

Adding 14 variables about the parents' age or age of death increased the R^2 from 0.158 to 0.185. This is roughly comparable to the increase from adding the self-assessed health variables. The coefficients on the other variables are little changed by adding the variables on the parents. For example, the difference in Plive75 associated with a difference in health status of "excellent' to "poor" is -.035, the same as it was when the parents' variables were excluded (Table 10).

The reference is someone whose parents died at age 65. The regressions have two types of variables: categorical variables, which are indicated in the table by a (1), and variables that are continuous in either the parents' age or age of death. The ages are normalized to be zero at 65. If the mother is alive (44% of the observations),

---

[8]We do not show coefficients on 32 other variables that were included in the regressions. The coefficients changed very little from what were reported in Tables 10 and 11.

Plive75 is predicted to be 0.073 greater (the coefficient on "mother alive" categorical variable) than if the mother had died at 65 , and it increases by 0.0014 for each year of the mother's age. Thus if she is alive and is 85, Plive75 is greater by 0.101 than if she had died at age 65. If the mother died before age 51 (7% of the observations), Plive75 is higher by 0.038 than if she had died at age 65. We imagine this is a reflection that early (accidental) deaths of parents do not affect probabilities of death of the child later in life. If the mother died between 51 and 64, Plive75 is almost the same as if her death had been at age 65. If the mother has died, her age at death increases Plive75 by 0.0039 per year, so that if she died at 85, Plive75 will be higher by 0.078 than if she had died at age 65. This is, of course, a rather large difference in Plive75: about two and a half years in life expectancy for men and almost four in the life expectancy for women. The other categorical variables cover missing data on parents' age and age at death, and while the coefficients can be large, the categories are not important in our sample.

The effects of the father's age or age at death are similar to those of the mother. For example if the father died at 85 rather than at age 65, Plive75 would be 0.092 greater.

It seems clear from these results that the respondents are aware that the age of their parents or the age of their parents' death has an influence on their own mortality risk and that they alter their reports on Plive75. The effects are large, particularly because the regressions control for self-assessed health, which is probably associated with the lifetime health status of the parents and their age at death.

The regression of Plive85 on the variables describing parents' age or age of death is similar. The R^2 increased from 0.134 to 0.168. As with Plive75 the health effects are about the same as when the parents' variables were excluded. The effects of disease are attenuated, and the effects of the parents' variables on Plive85 are similar to the effects on Plive75.

In these regressions no distinction was made between male and female respondents beyond a categorical variable for sex. Yet it is certainly plausible that males tend to form their expectations about longevity more from their father's age or age of death and females from their mother's age or age of death. To find if this is the case, we estimated the regression of Plive75 separately for each sex. We used the same set of 50 right-hand variables as in Table 12.

Table 13 shows just the coefficients on the variables describing the parents' age or age of death. Among females, Plive75 is increased by 0.082 if the mother is alive; yet by only 0.46 if the father is alive. Among males, Plive75 is increased by 0.068 if the father is alive and 0.047 if the mother is alive. This is just one example of the remarkable symmetry in the table: the coefficients on the mother's variables in the regressions over the data on females are about the same as the coefficients on the father's variables in the regressions over the data on males. For example, among women Plive75 increases by 0.0060 in the age of the mother's death; among men Plive75 increases by 0.0061 in the age of the father's death. In both cases the effects are considerably larger than the effects of the mother's age on the son's probability or the father's age on the daughter's probability.

This is summarized in Figures 11 and 12 which show the fitted values of Plive75 from the regressions. Among men Plive75 varies much more in the variables associated with the father than with the mother. Furthermore, the fitted probability is about the same whether the father reached his 80's and then died, or is still alive in his 80's. This is reasonable because of the high mortality risk among men in their 80's. The fitted values of Plive75 for women look almost like the fitted values for men except for the "mother" and "father" labels. This brings out rather clearly the symmetry of the coefficients in Table 13.

4. Probability of working

For studying the probabilities of working full-time past 62 or 65 (Pwork62 and Pwork65), we use the sample of full-time workers (hours of work greater than or equal to 35 per week) aged 51-61 because the transition from part-time work to full-time work is not common in this age group, and because part-time jobs have rather different characteristics from full-time jobs (Hurd, 1993).

We reported in section 2 the language of the questions we use to construct

Pwork62 and Pwork65. We rescaled the responses to the interval [0,1] and treated them as probabilities. Generally we will think of them as conditional probabilities: the probability of working full-time at 62 or 65 given working full-time at age t. It should be apparent, however, that the question has ambiguity: it could refer to working full-time anytime after the 62nd birthday or it could refer to working sometime after the respondent is no longer 62. As we will see, some respondents seemed to have the first interpretation and some the second.

It is much less straightforward to find population data to compare with Pwork62 and Pwork65 in the way we compared life table data with Plive75 and Plive85, but we will make two comparisons. The method behind the first is shown in Table 14. We have estimated the fraction of the population 55-59 working full-time and the fraction aged 63 working full-time from estimates of the fraction of full-time workers among all workers, and from labor force participation rates. According to this calculation the probability of working full-time at age 63 conditional on working full-time at ages 55-59 is $0.246 \div 0.537 = 0.457$. The average of Pwork62 over the 55-59 year-old full-time workers in the HRS is 0.478, which is remarkably close and which should increase our confidence that Pwork62 and Pwork65 measure conditional probabilities of working.

We can make an additional comparison based on the HRS data by using the observations of 62 and 63 year-old males who were interviewed but are not in the age-eligible population. They are husbands of age-eligible wives, and while they are not exactly representative of the 62 and 63 year-old male population (having to be married to younger women to be in the survey) we imagine they are sufficiently representative to give good estimates of the conditional probability of working full-time. We estimate the probability of working full-time at age 62 conditional on working full-time at t from the fraction of the HRS married males aged 62 who are working full-time and the fraction of the HRS married males working full-time at age t.

Figure 13 shows the average of Pwork62 of married males by age, and our estimates of the conditional probability of working full-time given full-time work at each of the ages 51-61. We show both the conditional probability of working to 62 and to 63 because of ambiguity in the HRS question: as discussed above, it is not clear whether the question refers to age 62 or age 63. Pwork62 is bounded by our estimates and has a modest but smaller upward trend. We conclude that the average of Pwork62 is reasonably close to population averages and to conditional probabilities calculated from the frequencies of full-time work in our sample.

Figure 14 has the distributions of Pwork62 and Pwork65. They have considerably less bunching at 0.5 than Plive75 and Plive85 and larger peaks at 0 and one. This is reasonable because respondents have considerable control over their work status, and many have decided either definitely to work past 62 or definitely not to. Some in the middle have not decided or they face greater random events than those who have.[9]

The conditional distribution of working (Pwork65 given Pwork62) is quite similar to the conditional distribution of Plive85: as in Figures 6-10, Pwork65 is on average less than Pwork62, and most respondents give smaller values for Pwork65 than for Pwork62. An interesting difference is in Figure 15, which is a graph of Pwork65 given that Pwork62 = 1.0. It shows that a substantial number of workers are certain of working past age 62 and certain of retiring before age 65. This corresponds to what we know about actual retirement: the retirement hazards are high at age 62 and very high at age 65. Many people plan to work until Social Security benefit eligibility (age 62) and then retire. Most of those who plan to work until 65 when they are eligible for full Social Security benefits, retire shortly after their 65 birthday.

Table 15 has our check for internal consistency. Unlike the case with Plive75 and Plive85, a probability of zero or one can be appropriate. The sum of the percentage giving those probabilities and of the percentage with Pwork62 > Pwork65 is 88.9%. Just 1.6% have Pwork62 < Pwork65.

The HRS asked workers if they had made plans or thought about retirement, and if so at what age did they plan to retire completely, change jobs, reduce hours, or become self-employed. If Pwork62 and Pwork65 are informed probabilities, we would expect that they would vary according to whether someone has thought about retirement.

---

[9]There are, of course, random events that affect retirement: health, layoff, and financial gains and losses, to name but several.

Figure 16 shows the fraction of workers that have not thought about retirement and the average of Pwork62 according to whether workers have or have not thought about retirement. The percentage that has not thought about retirement (37.7 percent over all ages) falls steadily with age from over 50 percent at age 51 to about 24 percent by age 61. Among those who have not thought about retirement, Pwork62 is rather steady except at age 61. At that age someone who has not thought about retirement and who therefore has no plans, probably has little choice but to continue to work. Accordingly Pwork62 increases to about 0.73.

It is likely that people do not think about retirement until several years before a possible retirement age. Therefore those who have thought about retirement are closer to their actual year of retirement. This means that workers in their early 50's who have thought about retirement will have small Pwork62. Workers in their late 50's who have thought about retirement will have higher Pwork62 because some plan to retire after age 62. The figure shows such an age pattern among those who have thought about retirement.

This view has the implication that the weighted average of Pwork62 (over those who have and have not thought about retirement) could be rather stable with age: until workers are in their late 50's the most important change in Pwork62 at the individual level is a decline that accompanies the switch from not having thought about retirement to having thought about it. The evolution in Pwork62 conditional on having thought about retirement could be rather minor. This means that the average variation by age does not represent the evolution in Pwork62 of an individual: it is the result of changing heterogeneity in the population with age.

It is well known that pension plans affect retirement. Defined benefit plans (DB), and to a lesser extent defined contribution plans (DC), affect retirement through the details of the structure of the plan. Typically DB plans have an age at which reduced pension benefits could be paid, and an age at which full benefits could be paid. Usually workers will not want to leave the firm a few years before those ages. Particularly after the age for full benefits, it often does not pay to remain with the firm, so many workers retire soon after qualifying for full benefits.

Table 16 has Pwork62 and Pwork65 classified by pension availability and by the details of the age of eligibility.[7] Having a DC plan reduces Pwork62 by about 0.04. However, if the earliest age for eligibility is greater than 62, Pwork62 increases to 0.64, which is considerably greater than Pwork62 of those with no pension. This effect is similar to the effect of Social Security on retirement at age 62, which is thought to act through a liquidity constraint.

Most workers with DB plans are eligible for full pensions benefits before the age of 62, and their Pwork62 is about 0.23 less than workers with no pension (Table 16). Among workers who must wait until 62 for a reduced or full benefit the probability increases by 0.11 to 0.41. If they must wait until age 65, their probability of working past age 62 increases to 0.67, which is greater than among workers with no pension. Thus, Pwork62 more than doubles as the age at which full benefits can be taken varies. The table shows similar variation in Pwork62 and Pwork65 as other details of the DB pension plan change. The variation is almost completely consistent with our knowledge of the effects of DB plans on actual retirement.

Figure 17 is based on an extract from Table 16. It shows among workers whose age for reduced benefits is less than 62 the variation in the probabilities of working as the age for full benefits varies. The effects are large, particularly when the age for full benefits increases from 62 to 65: apparently many workers plan to stay on the job past age 62 until they qualify for full benefits. If the age for full benefits is greater than 65, Pwork65 increases from 0.21 to 0.28; yet Pwork62 changes by very little. This illustrates the sensitivity of retirement plans to the details of DB pension plans.

5. Conclusion

Our criteria for judging the measures of subjective probabilities in the HRS were that they are good approximations to population probabilities, that they are internally consistent, and that they covary with other variables in the same way as in other data.

---

[7]We use information on the pension of the present job only.

On average the probabilities of living to 75 or 85 are close to averages in a lifetable from 1988. However, in view of the rapid change in mortality rates, it is not really clear how close they should be because we do not know how cohorts form their views about mortality risk.

Labor force participation rates have been stable for a number of years, implying that retirement behavior has been stable. Therefore, the mean of Pwork62 should be close to the population average, at least when compared with the difference between Plive75 and a population probability from a life table. Indeed, the HRS measure of the subjective probability of working full-time past 62 is within several percentage points of a population estimate.

The subjective probabilities are in general internally consistent. To the extent that they are not, an analyst should model the process that causes the inconsistency. The process surely includes observation error, and in this regard is no different from almost all economic variables. Usually, however, the respondent and the analyst share a common understanding of the meaning of a survey question. This is undoubtedly not always true for the questions about subjective probabilities, and that difference needs to be taken into account.

The sharpest test of the subjective probabilities comes from their covariation with other variables. The probabilities of living to 75 or 85 vary in a systematic and reasonable way with diseases, socio-economic status, self-assessed health, and indicators of family longevity. The probabilities of working past 62 or 65 vary with personal, financial and job variables in ways that are consistent with what has been found in other data. On average, therefore, the subjective probabilities will correctly predict some of the variation in outcomes. For example, workers with defined benefit pensions will retire earlier than workers without pensions, and because workers with defined benefit plans have lower subjective probabilities of working past 62, small values of Pwork62 will correctly predict early retirement. Of course, what we hope is that conditional on observable characteristics, the subjective probabilities will be good predictors of retirement, which will allow us to observe and control for individual heterogeneity. Finding whether this happens will require observations in the panel data. From the cross-section, however, we conclude that the measures of subjective probabilities in the HRS show great promise for making a substantial contribution to our understanding of decision making under uncertainty.

Table 1
Average probabilities of living to 75 or 85

	Men		Women		All	
	Age 75	Age 85	Age 75	Age 85	Age 75	Age 85
HRS data*	0.62	0.39	0.66	0.46	0.65	0.43
1988 life table*	0.59	0.24	0.75	0.43	0.68	0.34
From age 55						
HRS data	0.64	0.40	0.67	0.46	0.66	0.43
1980 life table	0.54	0.21	0.73	0.41	0.64	0.31
1988 life table	0.59	0.24	0.74	0.43	0.68	0.33
2000 life table	0.62	0.28	0.78	0.51	0.70	0.40

Source: Authors' calculations from HRS and various life tables for the U.S.
*Ages 51-61 only.

Table 2
Comparison of probabilities of living to 75 and 85

Probability comparison	Percent of respondents
Plive75 > Plive85	70.1
Both probabilities = 0	6.9
Both probabilities = 0.5	4.7
Both probabilities = 1.0	9.2
Both probabilities = some other value	6.6
Plive75 < Plive85	2.5

Source: Authors' calculations from HRS.

Table 3
Probability of living to 75 or 85: Income and Wealth

Quartile	To 75		To 85	
	Income	Wealth	Income	Wealth
first	0.59	0.57	0.39	0.39
second	0.63	0.62	0.42	0.40
third	0.66	0.66	0.43	0.44
fourth	0.70	0.70	0.48	0.47

Source: Authors' calculations from HRS.

Table 4
Probability of living to 75 or 85: Education

Education level	Observations	Past 75	Past 85
Less than high school	2190	0.57	0.37
High school	2855	0.65	0.42
Greater than high school	2896	0.69	0.48

Source: Authors' calculations from HRS.

Table 5
Probability of living to 75 or 85: Self-assessed Health Status

	Men			Women		
Health status	Observations	75	85	Observations	75	85
Excellent	793	0.75	0.53	1006	0.78	0.58
Very good	998	0.68	0.42	1236	0.71	0.50
Good	1037	0.61	0.37	1162	0.64	0.44
Fair	449	0.47	0.27	645	0.53	0.33
Poor	286	0.34	0.16	328	0.40	0.23

Source: Authors' calculations from HRS.

Table 6
Probability of living to 75 or 85: Smoking Status

Smoking status	Observations	Age 75	Age 85
Never smoked	2927	0.67	0.47
Not now	2878	0.65	0.43
Yes	2138	0.60	0.38

Source: Authors' calculations from HRS.

Table 7
Probability of living to 75 or 85: Drinking

Drinks per day	Observations	Age 75	Age 85
Doesn't drink	3126	0.61	0.41
less than 1	3593	0.67	0.45
1-2	812	0.68	0.44
3-4	295	0.60	0.36
5 or more	112	0.55	0.33

Source: Authors' calculations from HRS.

Table 8
Probability of living to 75 or 85: Health Status and Education

	Education		
Health status	Less than high school	High school	More than high school
	Living to 75		
Excellent	0.71	0.77	0.78
Very good	0.68	0.69	0.71
Good	0.60	0.63	0.64
Fair	0.51	0.48	0.52
Poor	0.36	0.34	0.44
	Living to 85		
Excellent	0.52	0.53	0.58
Very good	0.45	0.45	0.48
Good	0.40	0.40	0.41
Fair	0.33	0.26	0.32
Poor	0.19	0.17	0.25

Source: Authors' calculations from HRS.

Table 9
Probability of living to 75 or 85: Health status and drinking

	Health Status				
Drinks per day	Excellent	Very good	Good	Fair	Poor
		to 75			
Don't drink	0.74	0.69	0.62	0.50	0.37
< 1	0.77	0.70	0.64	0.51	0.40
1-2	0.80	0.71	0.62	0.50	0.24
3-4	0.74	0.65	0.56	0.44	0.40
5 +	0.71	0.59	0.54	0.59	0.30
		to 85			
Don't drink	0.54	0.46	0.43	0.32	0.18
< 1	0.57	0.47	0.40	0.29	0.24
1-2	0.54	0.46	0.37	0.27	0.13
3-4	0.48	0.43	0.30	0.22	0.21
5+	0.38	0.39	0.36	0.38	0.15

Source: Authors' calculations from HRS

	No health variables	Health variables included	
	parameter	parameter	S.E.
Drinks <1 per day	0.019*	0.009	0.008
Drinks 1-2	0.026*	0.018	0.013
Drinks 3-4	-0.007	-0.015	0.019
Drinks 5+	0.003	0.010	0.032
Education < 12	-0.046*	-0.013	0.009
Education > 12	0.021*	0.013	0.008
Ever high blood pressure	-0.034*	-0.009	0.008
Ever diabetes/high blood sugar	-0.037*	0.009	0.012
Cancer/malignant tumor	-0.072*	-0.040*	0.016
Chronic lung disease	-0.058*	-0.005	0.014
Ever heart problems	-0.069*	-0.030*	0.013
Angina/chest pains	-0.062*	-0.025	0.022
Congestive heart failure	-0.063*	-0.018	0.030
Ever had stroke	-0.023	0.015	0.022
Arthritis/Rheumatism	-0.029*	0.002	0.008
Weight (100 lbs)	0.011	0.019	0.011

Source: Authors' calculations from HRS
Note: Average of P(75) = 0.649 based on 6095 observations. $R^2 = 0.158$
*Significant at 5% level

Table 10
Determinants of Probability of living to 75: Self-assessed Health Status

	No health variables	Health variables included	
Variable	Parameter	Parameter	Standard error
Intercept	0.745*	0.791*	0.030
Household income (100 thousand)	0.015	0.003	0.009
Wealth (millions)	0.016*	0.003	0.001
Age	0.004*	0.004*	0.001
Married	0.001	0.000	0.020
Male	-0.049*	-0.048*	0.009
Light phys. activity: 1-2 per week	-0.021*	-0.021*	0.009
1-3 per month	-0.017	-0.010	0.013
< 4 per month	-0.009	-0.005	0.014
never	-0.041*	-0.013	0.014
Heavy phys. activity: 1-2 per week	-0.004	0.003	0.014
1-3 per month	-0.005	0.002	0.015
< 4 per month	-0.040*	-0.031*	0.012
never	-0.058*	-0.032*	0.011
Health: Very good		-0.057*	0.010
good		-0.122*	0.010
fair		-0.232*	0.014
poor		-0.345*	0.019
Race (White=1)	-0.040*	-0.052*	0.010
Formerly smoked	0.001	0.001	0.008
Currently smokes	-0.037*	-0.026*	0.009

Table 11
Determinants of Probability of living to 85: Self-assessed Health

Variable	No health variables	Health variables included	
	Parameter	Parameter	Standard error
Intercept	0.592*	0.644*	0.033
Household income (ten thousand)	0.015	0.004	0.010
Wealth (millions)	0.005	0.002	0.009
Age	0.004*	0.004*	0.001
Married	-0.028	-0.029	0.022
Male	-0.084*	-0.083*	0.010
Light phys. activity: 1-2 per week	-0.034*	-0.033*	0.010
1-3 per month	-0.023	-0.016	0.014
< 4 per month	-0.007	-0.003	0.016
never	-0.047*	-0.022	0.015
Heavy phys. activity: 1-2 per week	-0.004	0.003	0.016
1-3 per month	-0.011	-0.002	0.017
< 4 per month	-0.032*	-0.021	0.014
never	-0.061*	-0.036*	0.013
White	-0.084*	-0.096*	0.011
Health: Very good		-0.076*	0.011
good		-0.129*	0.011
fair		-0.228*	0.015
poor		-0.321*	0.021
Formerly smoked	-0.006	-0.007	0.009
Currently smokes	-0.041*	-0.030*	0.011

	No health variables	Health variables included	
	Parameter	Parameter	Standard error
Drinks < 1 per day	0.010	0.001	0.009
Drinks 1-2	0.010	0.002	0.014
Drinks 3-4	-0.011	-0.018	0.021
Drinks 5+	0.015	0.020	0.036
Education < 12	-0.018	0.012	0.010
Education > 12	0.038*	0.030*	0.009
Ever high blood pressure	-0.041*	-0.017*	0.009
Ever diabetes/high blood sugar	-0.056*	-0.015	0.014
Cancer/malignant tumor	-0.041*	-0.012	0.018
Chronic lung disease	-0.053*	-0.005	0.015
Ever heart problems	-0.097*	-0.060*	0.014
Angina/chest pains	-0.037	-0.006	0.024
Congestive heart failure	0.016	0.054	0.033
Ever had stroke	0.028	0.061*	0.025
Arthritis/Rheumatism	-0.040*	-0.010	0.008
Weight (100 lbs)	0.029*	0.037*	0.012

Source: Authors' calculations from HRS
Note: *Significant at 5% level.
Note: Average of P(85) = 0.432 based on 6077 observations. $R^2 = 0.134$

Table 12
Effect of Parents' Age or Age of Death on Probability of living to 75

Variable	Coefficient	Standard error
(32 additional coefficients not listed in this table)		
Health very good	-0.060	0.010
good	-0.124	0.010
fair	-0.233	0.014
poor	-0.347	0.019
Mother alive (1) (44%)	0.073	0.018
Mother's age-65 if alive	0.0014	0.001
Mother alive, age missing (1) (0.3%)	-0.089	0.072
Mother's age at death < 51 (1) (7%)	0.038	0.017
Mother's age at death 51-64 (1) (10.5%)	-0.003	0.015
Mother's age at death-65 if gt 65 (35.2%)	0.0039	0.001
Mother dead, age missing (1) (3%)	0.066	0.025
Father alive (1) (18%)	0.053	0.025
Father's age-65 if alive	0.0010	0.001
Father alive, age missing (1) (0.2%)	0.152	0.089
Father's age at death < 51 (1) (9%)	0.028	0.014
Father's age at death 52-64 (1) (18%)	0.007	0.012
Father's age at death-65 if gt 65 (49.8%)	0.0046	0.001
Father dead, age missing (1) (5%)	0.066	0.019

Source: Authors' calculations from HRS. $R^2 = 0.185$.

Table 13
Summary of effect of parents' age or age of death on probability of living to 75 or 85

	Living to 75		Living to 85	
	Females	Males	Females	Males
Parent alive				
mother	0.082*	0.047	0.079	0.044
father	0.046	0.068	0.070	0.067
mother's age - 65	0.0024	0.0018	0.0060*	0.0026
father's age - 65	-.0003	0.0019	0.0009	0.0030
Parent dead				
M-age < 51	0.047	0.044	0.089*	0.026
F-age < 51	0.007	0.047*	0.009	0.042
50 < M-age < 65	-.004	0.009	0.034	0.032
50 < F-age < 65	0.001	0.008	0.001	0.012
M-age - 65	0.0060*	0.0024*	0.0080*	0.0034*
F-age - 65	0.0039*	0.0061*	0.0041*	0.0061*

Source: Authors' calculations from HRS.
Note: M-age = Mother's age at death; F-age = Father's age at death.
Note: Extract from regressions with 51 right-hand variables
*Significant at 5% level

Table 14
Calculation of conditional probability of working full-time at age 63

Age	Fraction of workers working full-time (1987)	Labor force participation rate (1988-1989)	Fraction of population working full-time
55-59	0.805	0.668	0.537
63-65	0.669	0.367*	0.246*

Sources: Sum and Fogg (1990), and CPS
*63 year-olds

Table 15
Comparison of probabilities of working past 62 and 65

Probability comparison	Percent of respondents
P62 > P65	54.8
Both probabilities = 0	28.3
Both probabilities = 0.5	2.9
Both probabilities = 1.0	5.8
Both probabilities = some other value	6.5
P62 < P65	1.6

Source: Authors' calculations from HRS.

Table 16
Probabilities of working past 62 and 65: Pension effects

Explanation		NOBS	Past 62	Past 65
No plan		924	0.53	0.31
Defined contribution		764	0.49	0.24
Earliest age >62		63	0.64	0.37
Earliest age missing		379	0.48	0.24
Defined benefit				
age for early benefits	age for full benefits			
less than 62	less than 62	571	0.30	0.14
less than 62	62	167	0.31	0.09
62	62	194	0.41	0.18
less than 62	65	165	0.55	0.21
62	65	117	0.58	0.22
65	65	80	0.67	0.29
less than 62	greater than 65	21	0.57	0.28
greater than 65	greater than 65	11	0.64	0.44

Source: Authors' calculations from the HRS.

Figure 1
Distribution of probabilities of living

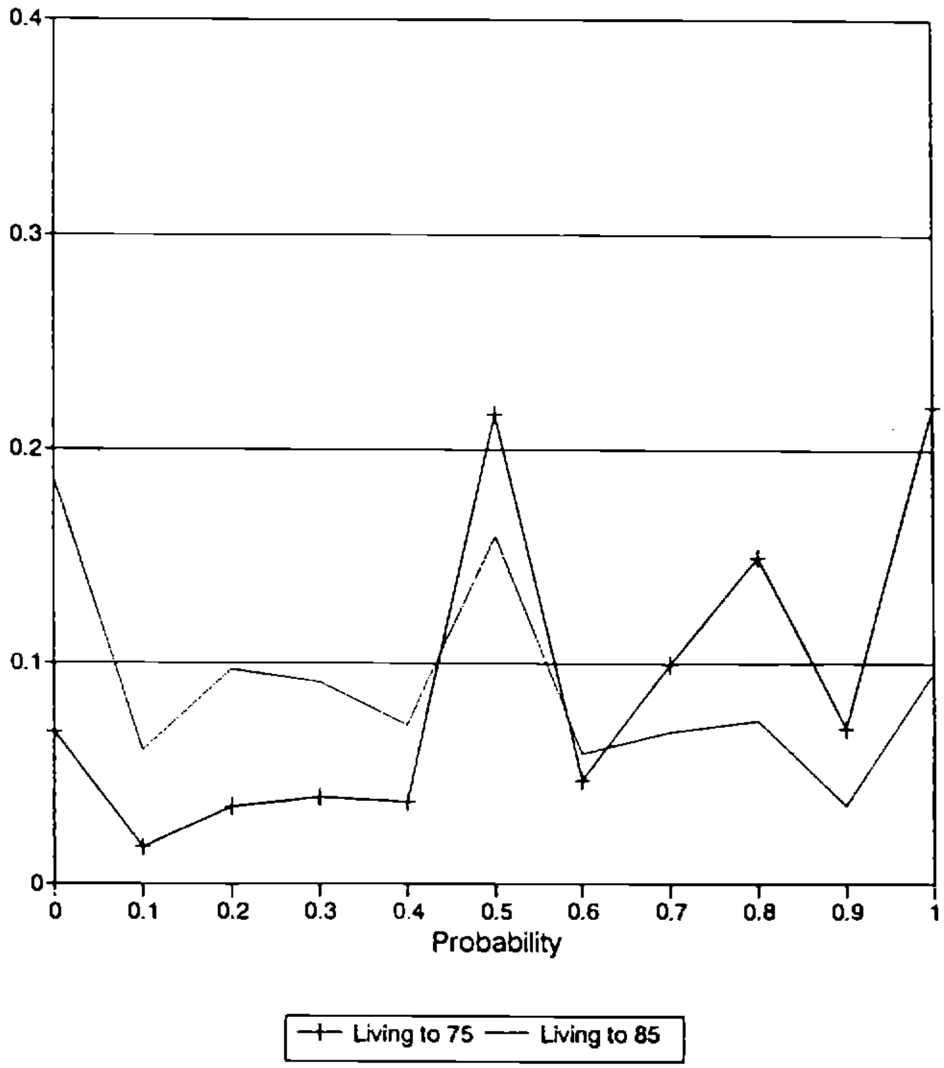


Figure 2
Probability of living: men

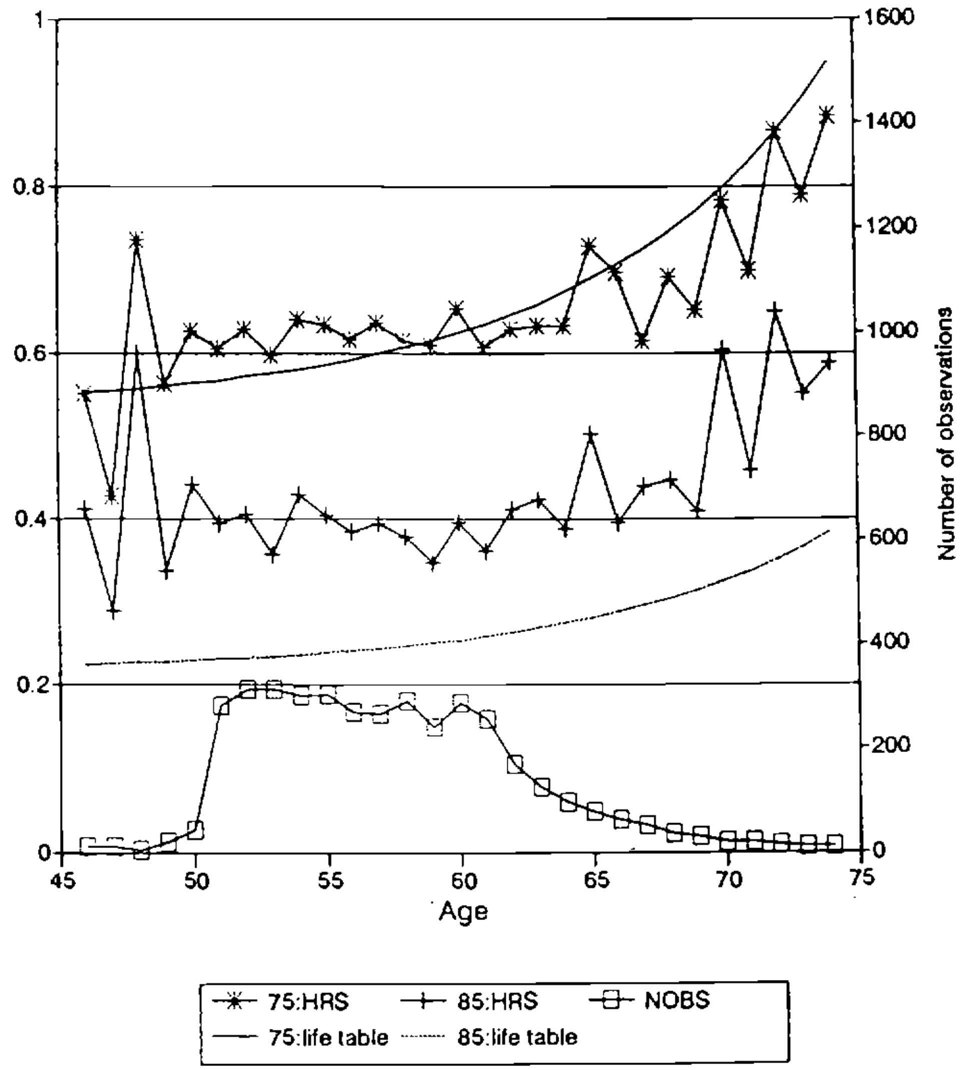


Figure 3
Probability of living: women

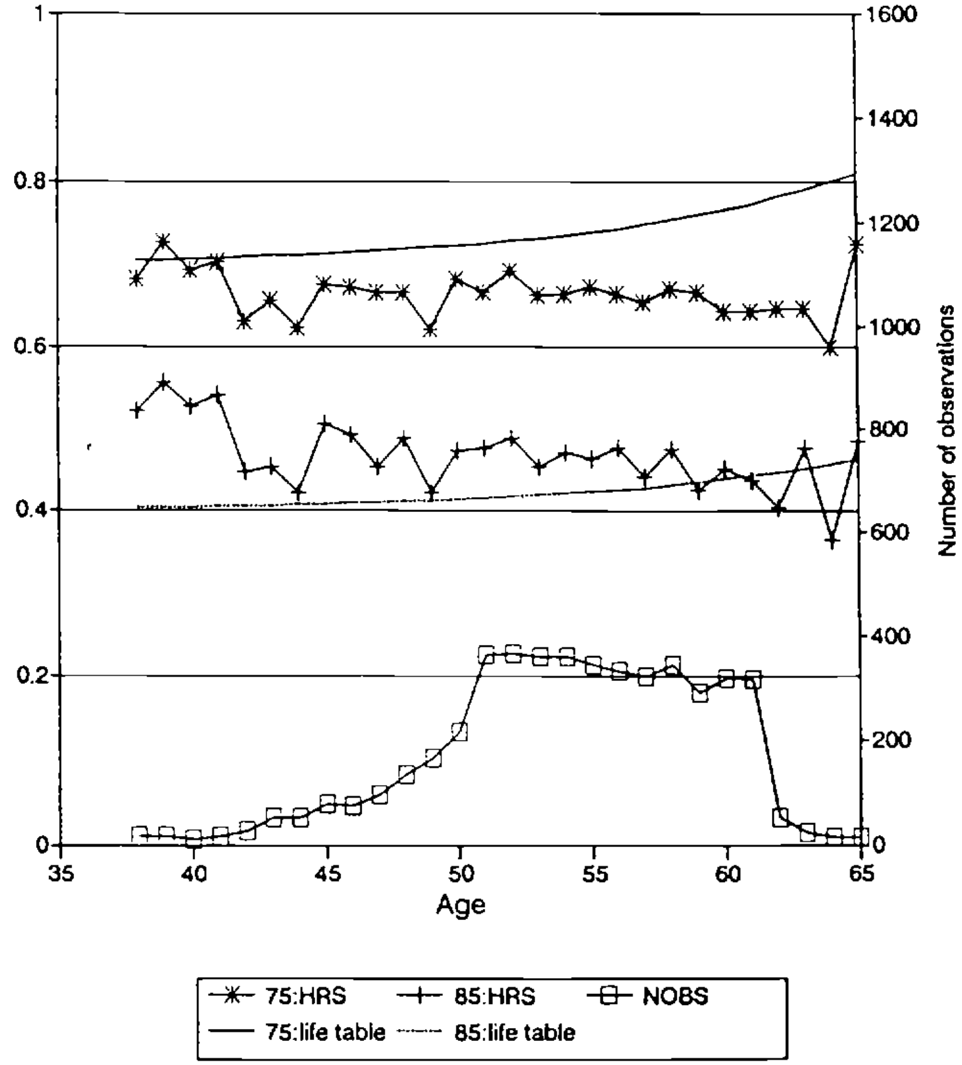


Figure 4
Prob. of men living to 75: lifetables

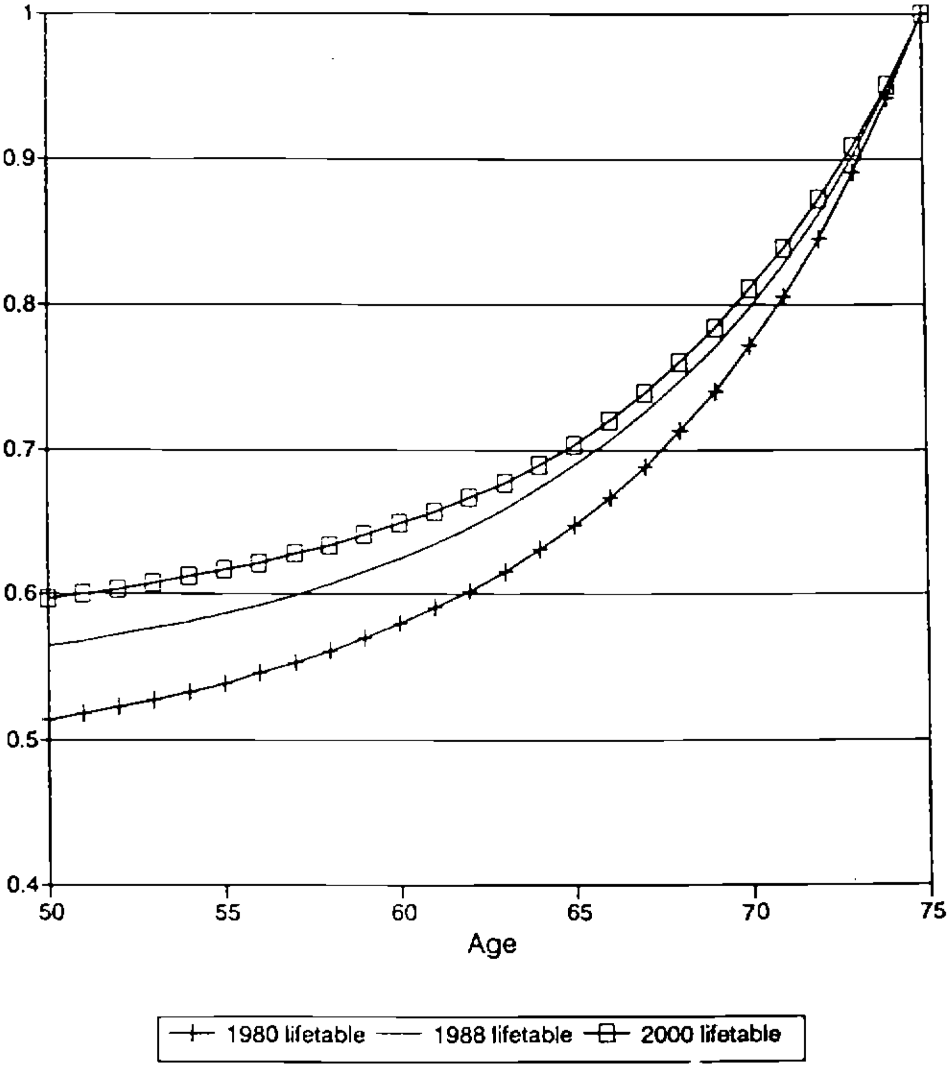


Figure 5
Prob. of men living to 85: lifetables

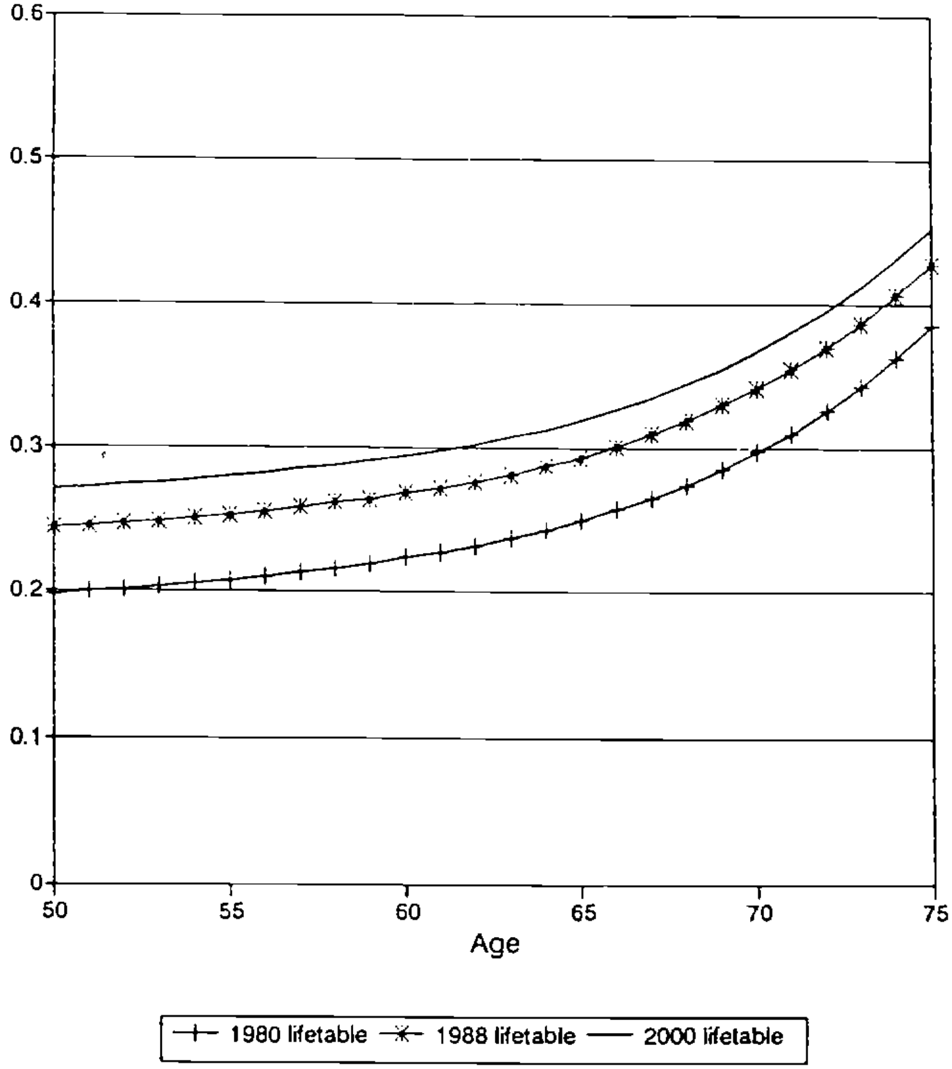


Figure 6
Prob. of living to 85 given prob. of 75

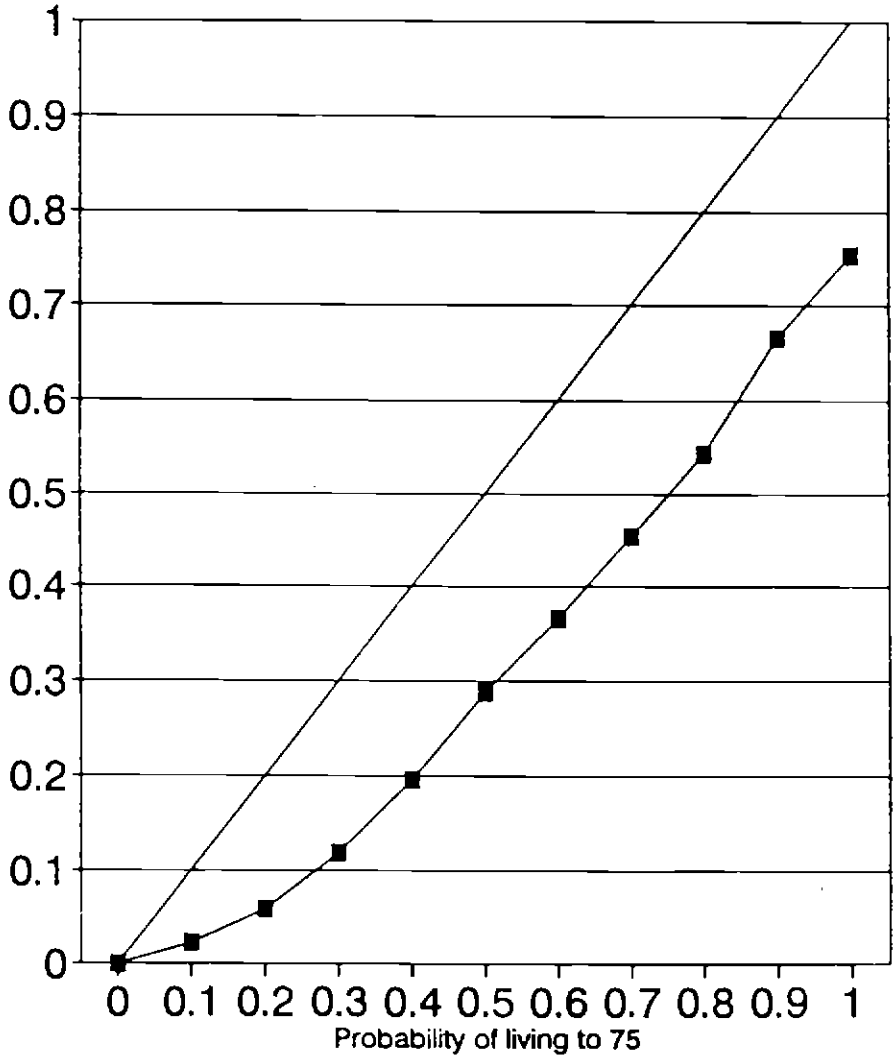


Figure 7
Distribution of P85 given P75 = 0.2

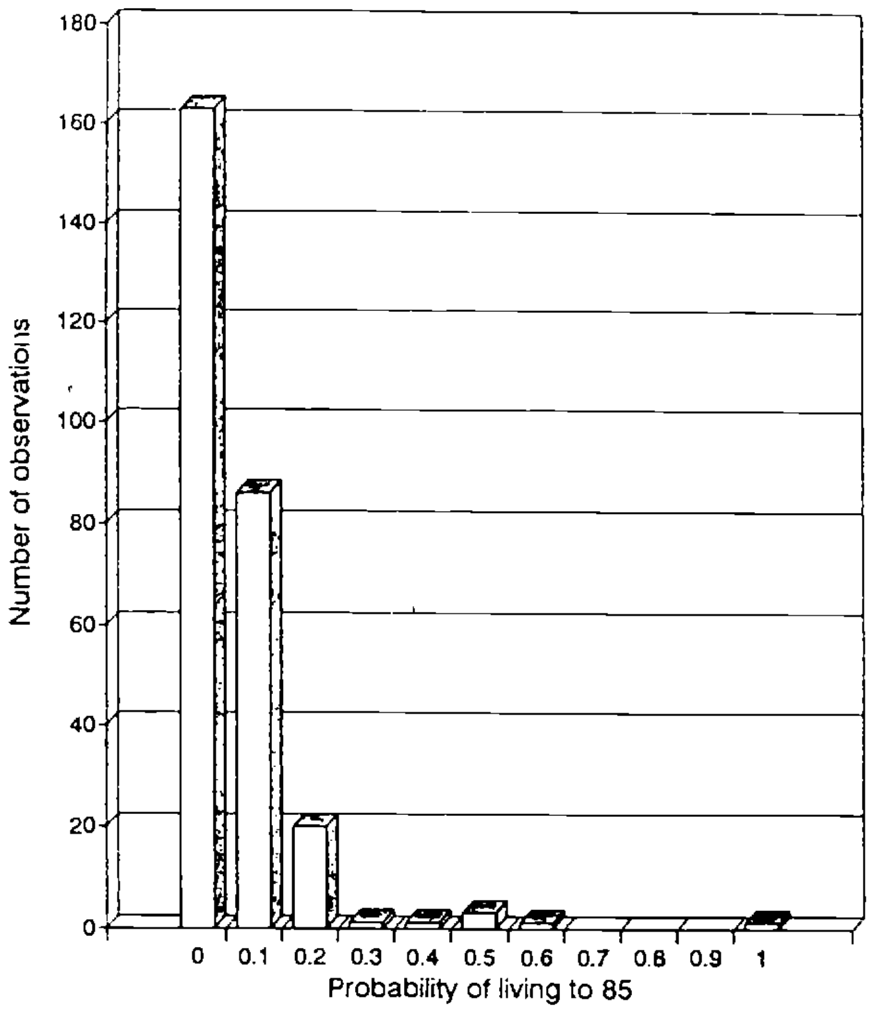


Figure 8
Distribution of P85 given P75 = 0.5

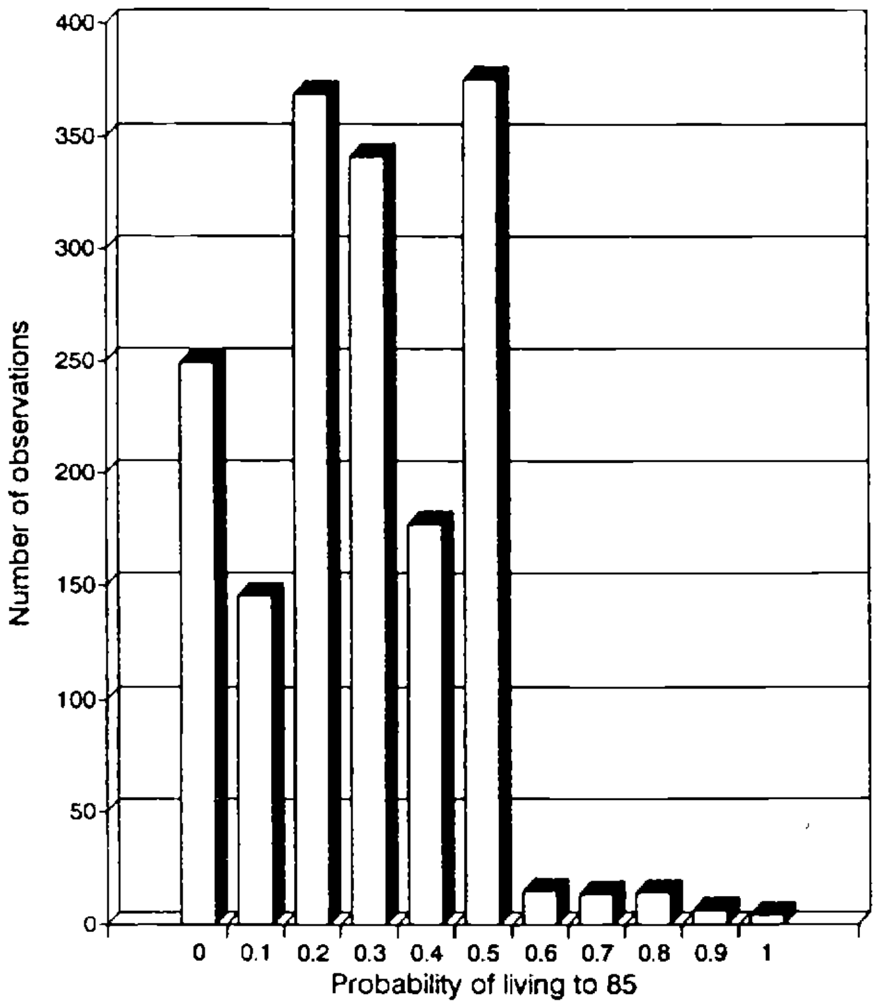


Figure 9
Distribution of P85 given P75 = 0.8

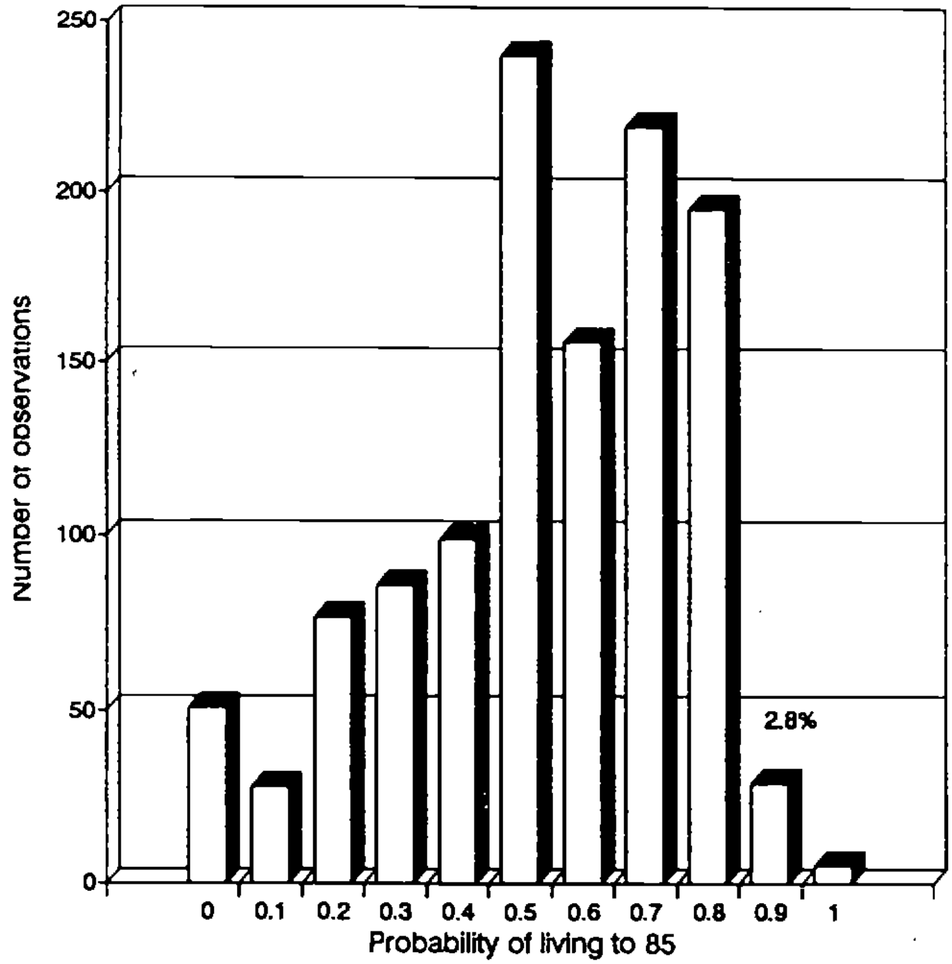


Figure 10
Distribution of P85 given P75 = 1.0

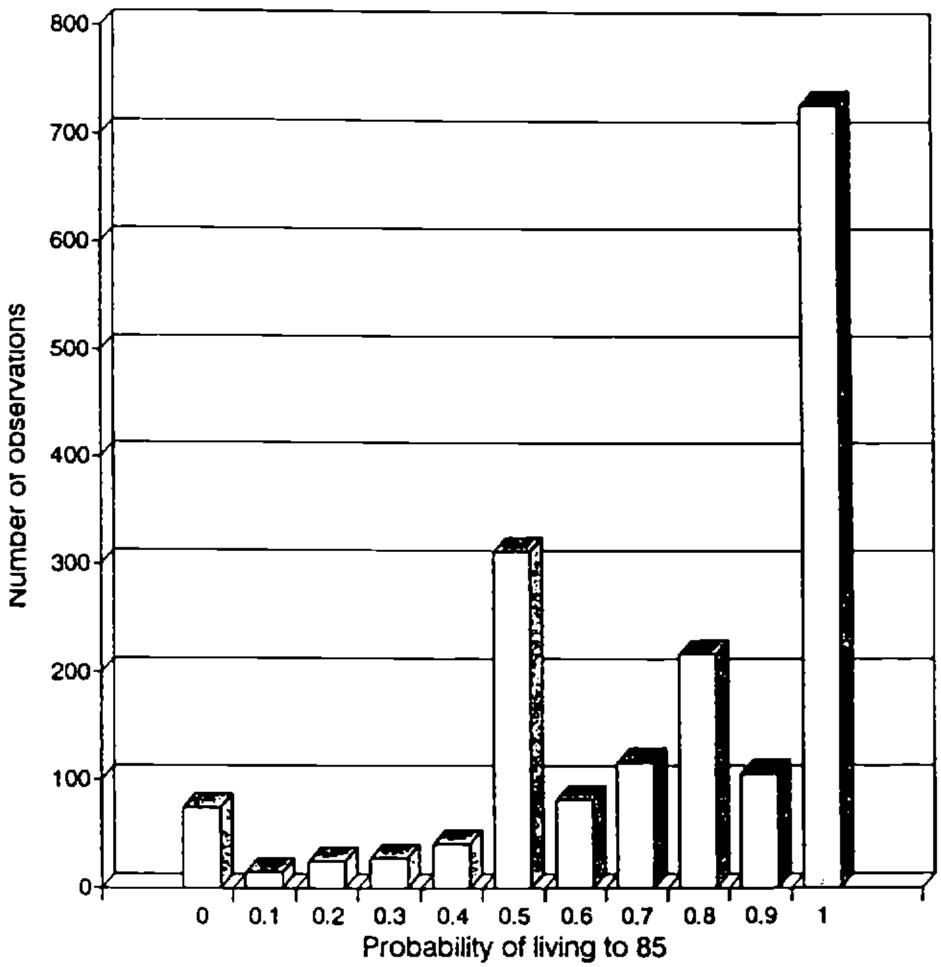


Figure 11
Probability of men living to 75

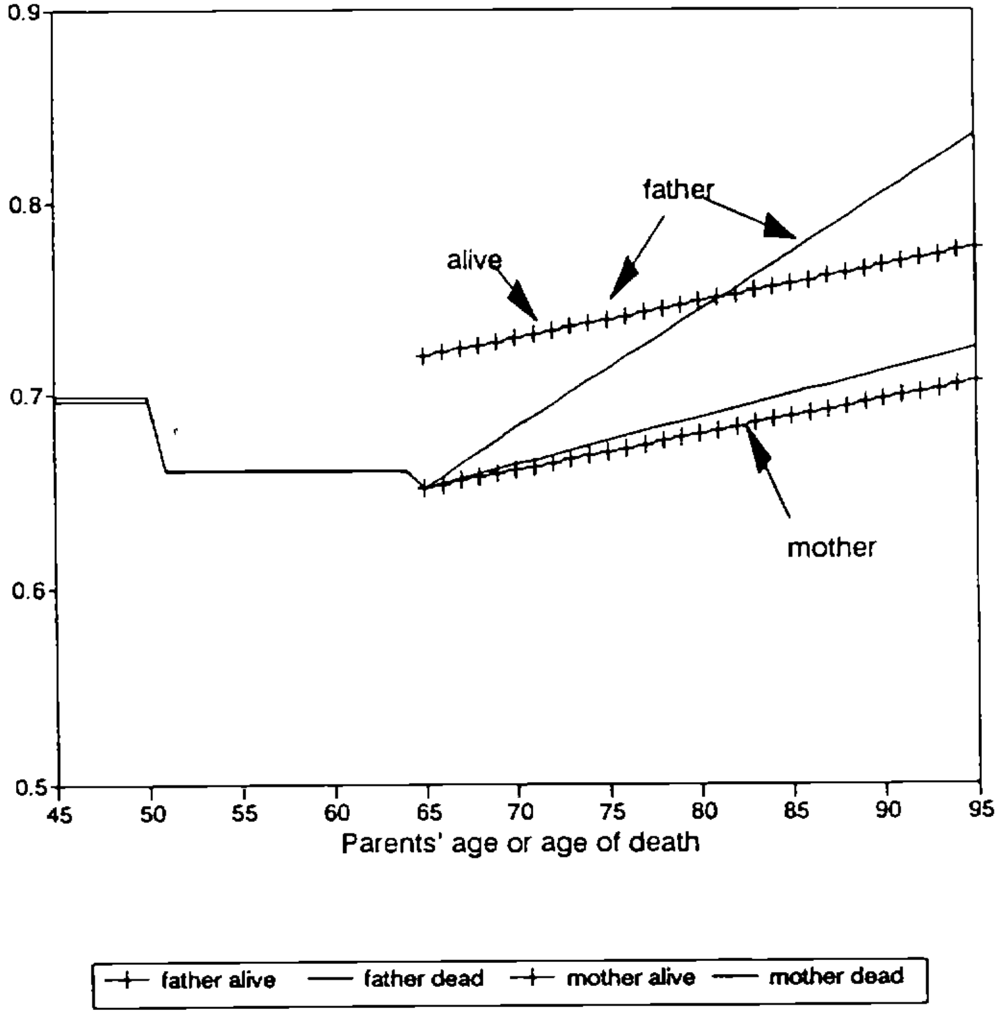


Figure 12
Probability of women living to 75

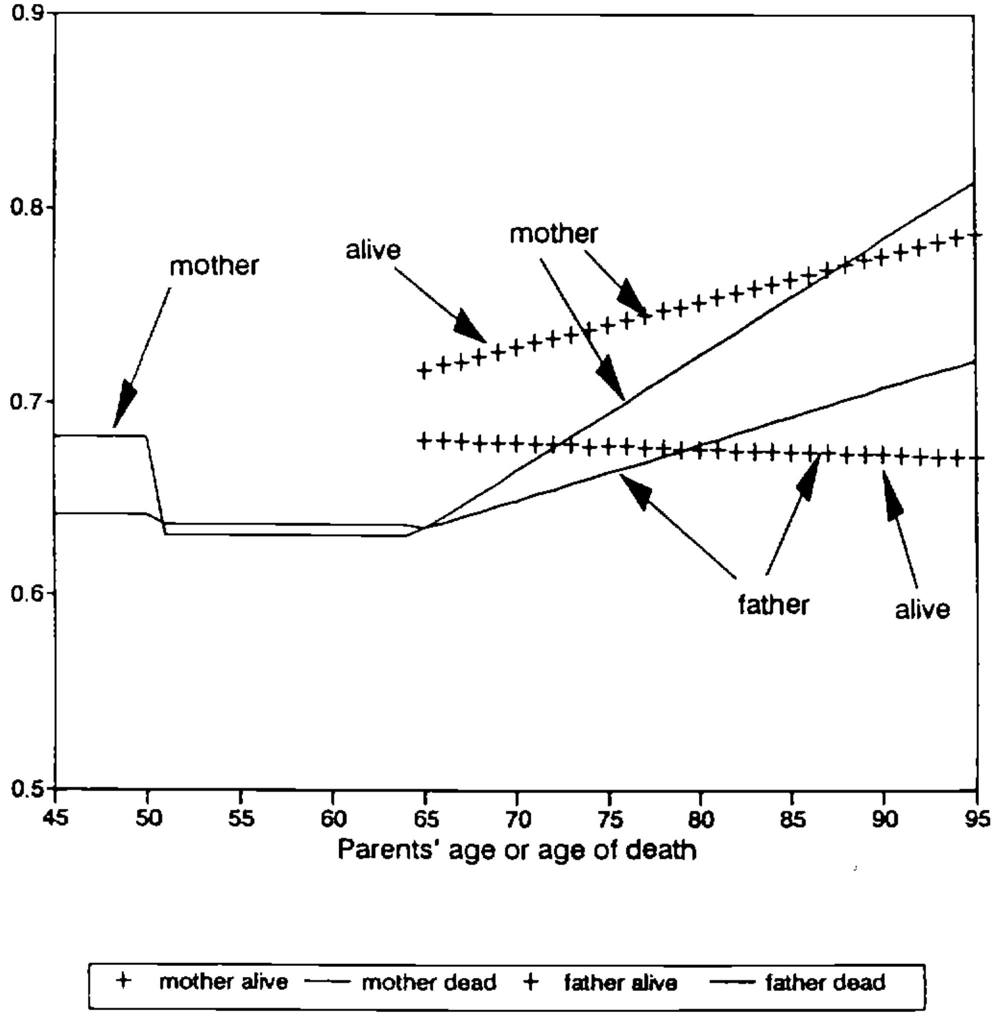


Figure 13
Conditional prob. of working past 62

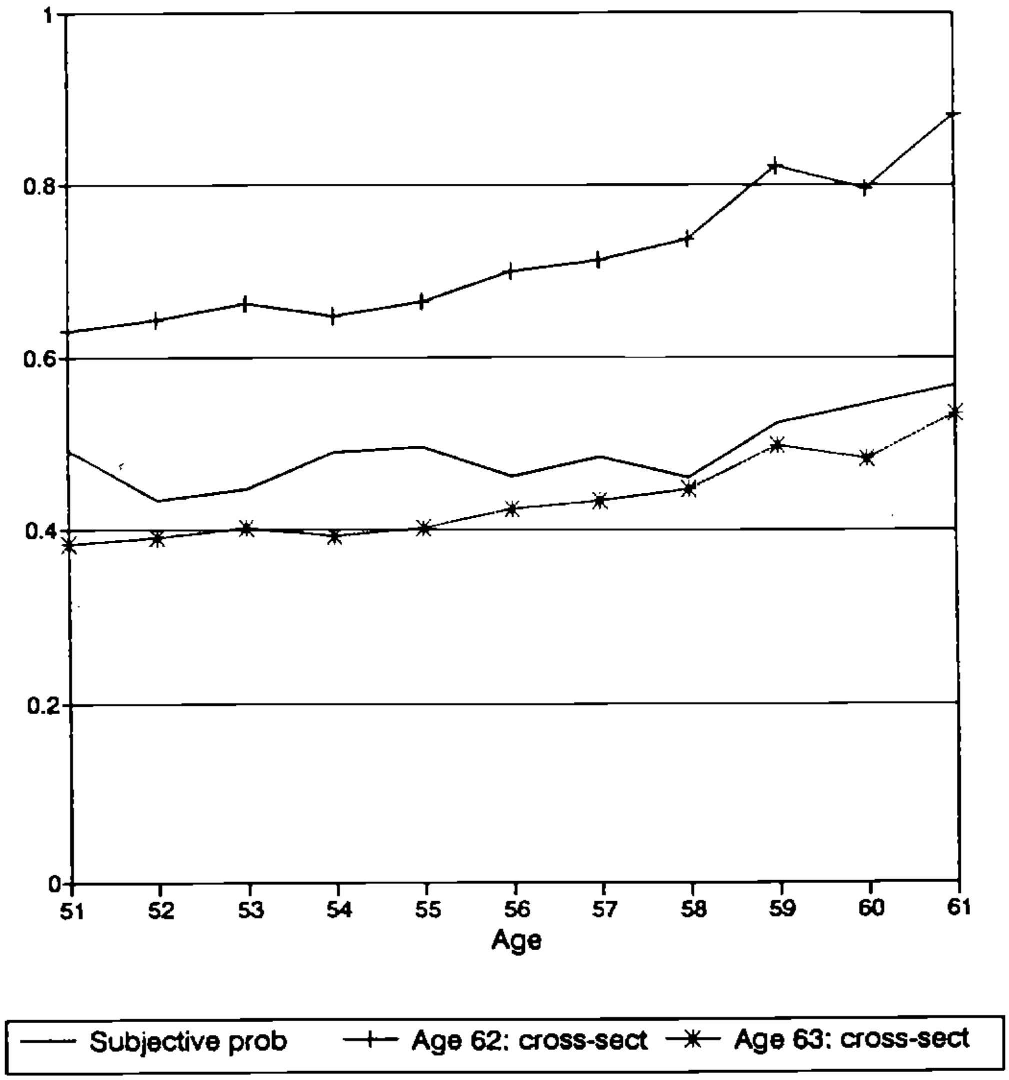


Figure 14
Distribution of probs. of working

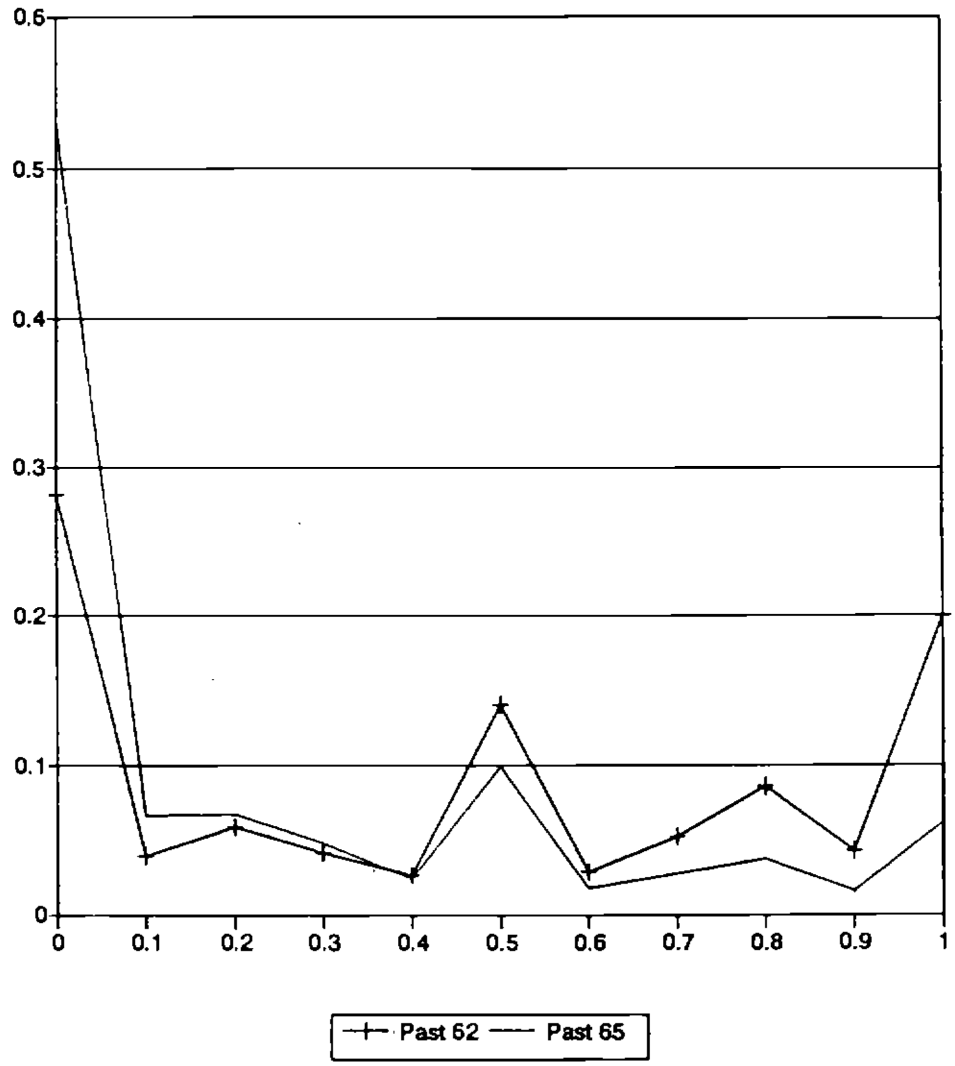


Figure 15
Distn. of Pwork65 given Pwork 62 = 1.0

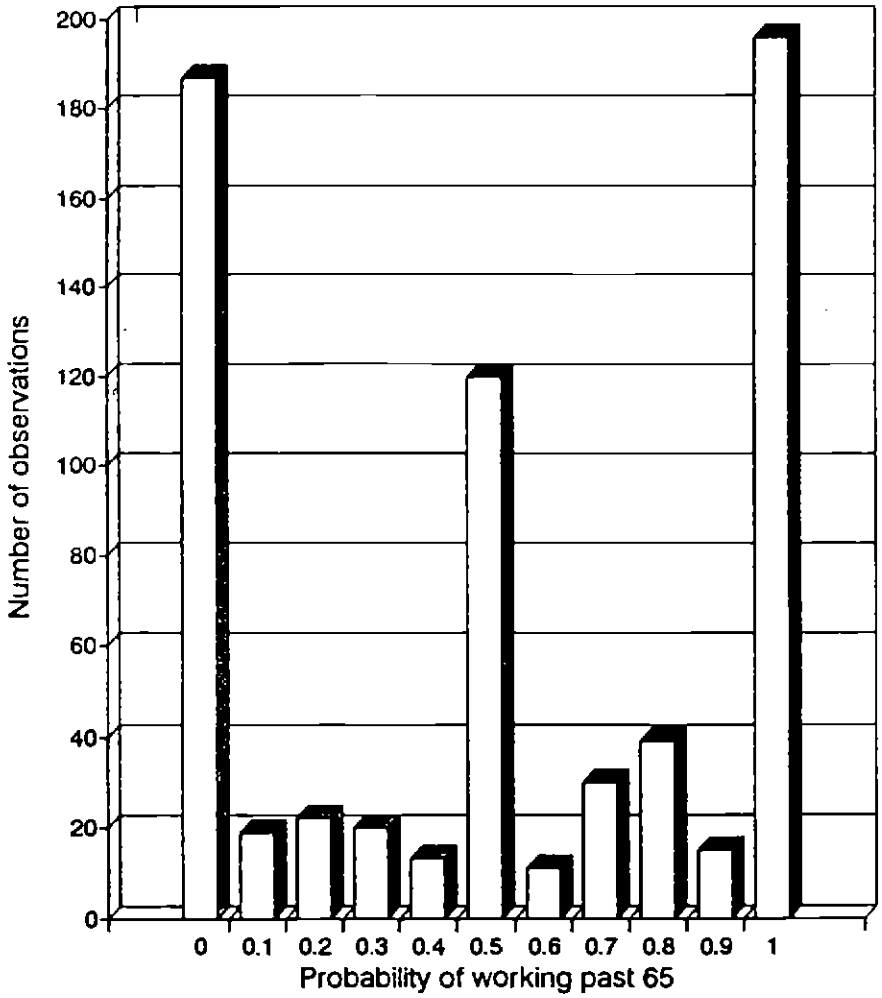


Figure 16
Probability of working past 62

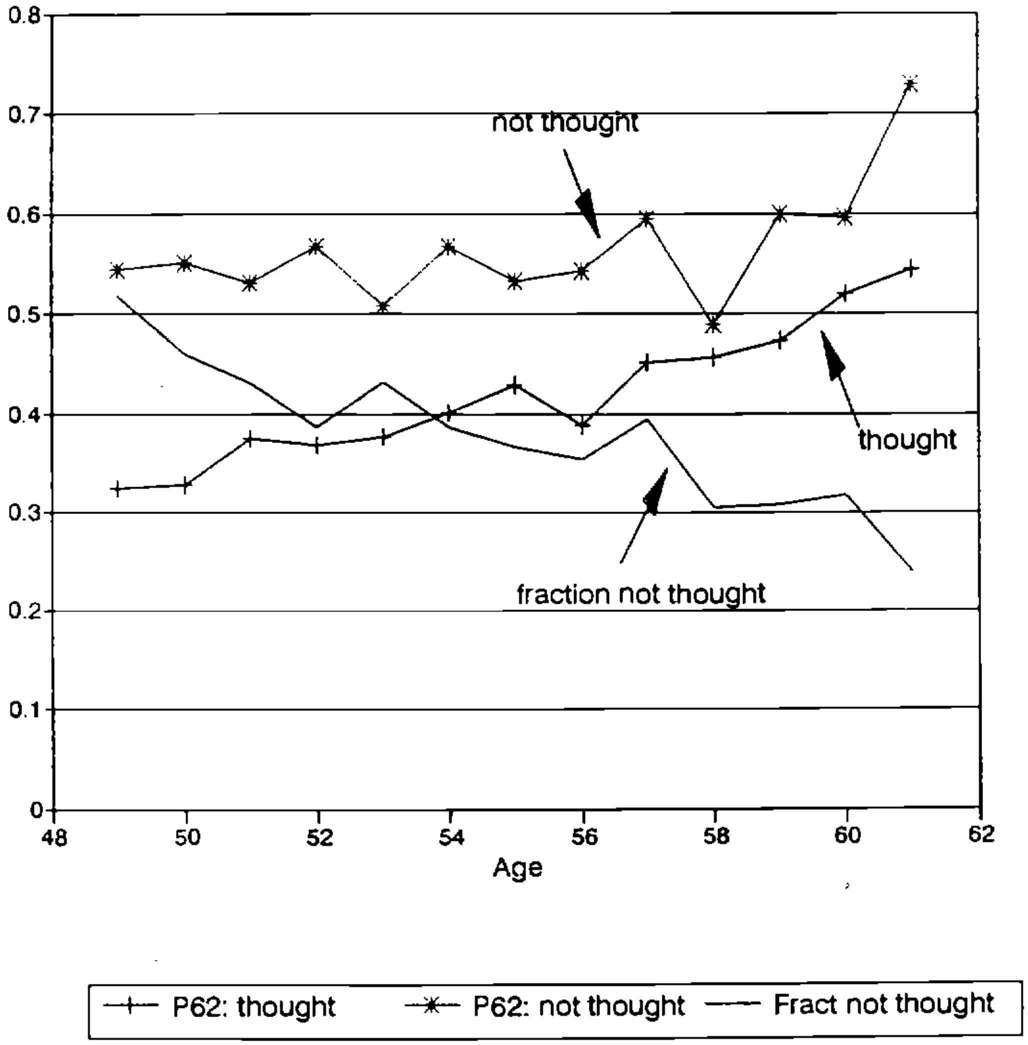
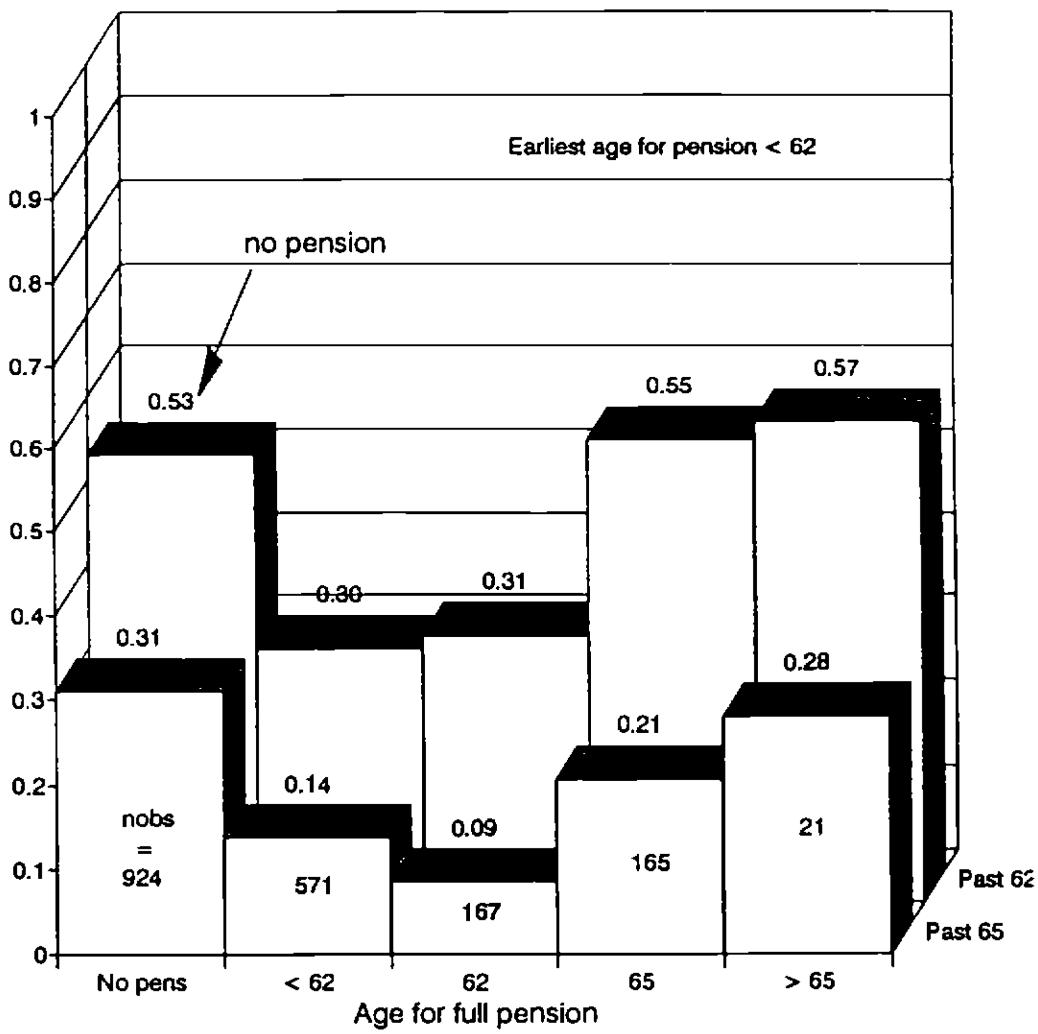


Figure 17
Probability of working past 62 or 65



References

Hurd, Michael D. (1993) "The Effect of Labor Market Rigidities on the Labor Force Behavior of Older Workers," NBER Working Paper 4462.

Juster, F. Thomas and Richard Suzman (1993) "The Health and Retirement Study: An Overview," presented at the HRS Early Results Workshop, University of Michigan, September.