DRAFT: DO NOT QUOTE

The Empirical Relationship between Lifetime Earnings and Mortality

Julian Cristia
Congressional Budget Office
julian.cristia@cbo.gov

February 2007

Abstract

Researchers have aimed to estimate the extent of differential mortality across socioeconomic groups by classifying individuals using income in the previous year. The first problem with this strategy is reverse causation: the same health shocks that lead to increase morbidity also cause income to decline. Second, annual income is a noisy measure of permanent income. This paper aims to tackle these two drawbacks by classifying individuals using measures of lifetime earnings. These measures are computed as long averages of annual earnings for ages when individuals have strong attachment to the labor market. A unique data set constructed matching the Survey of Income and Program Participation to Social Security records on earnings and mortality is used. Results indicate that the gradient between mortality and lifetime earnings is very strong (e.g. men aged 35-49 in the bottom quintile of lifetime earnings have 6.2 times higher one-year age-adjusted mortality than those in the top quintile). It is also found that the gradient is smaller for women than for men, varies when own versus household earnings is used and decreases with age. Adjusting for race, marital status and education only slightly decreases the estimated differentials. Finally, there is evidence that differential mortality by lifetime earnings has increased in the last twenty years.

The analysis and conclusions expressed in this paper are those of the author and should not be interpreted as those of the Congressional Budget Office. The author wishes to thank comments and suggestions by Howard Iams, John Sabelhaus and Hilary Waldron.

I. Introduction

Differential mortality by race, ethnicity, education, marital status, and economic measures has been extensively investigated in the literature (see a survey by Feinstein, 1993). In particular, researchers have aimed to estimate differential mortality across socioeconomic groups by measuring differences in mortality rates in a given year across groups defined by income in the previous year (Kitagawa and Hauser, 1973; Duleep, 1989; Sorlie, Backlund and Keller, 1995). This strategy faces two drawbacks. First, this approach suffers from reverse causation: individuals who experience health shocks (which increase their mortality probability) may drop out from the labor market and simultaneously suffer a drop in income. As a result, this approach will overstate the true correlation between permanent income and mortality. Second, yearly income is a noisy measure of permanent income. Taking into account only this effect, we should expect that estimates of differential mortality by income in a specific year will underestimate the extent of differential mortality by permanent income.[1]

In this paper, I aim to tackle the two aforementioned drawbacks by classifying individuals using measures of lifetime earnings. These measures are constructed using long averages of past earnings. For individuals older than 50, earnings while aged 38 to 47 are used to capture the years when the person was closest attached to the labor market. For younger individuals, averages ranging from 5 to 10 years were computed without including the immediate preceding 3 years (e.g. for individuals aged 40, earnings while aged 33 to 37 are used). In this way, the problem of reverse causation is at least partially

[1] Evans and Singleton (2006) explore how large this effect is by comparing the correlation between earnings in one year and mortality to the correlation of annual earnings averages of varying length to mortality.

addressed by dropping, in the computation of the earnings average, years immediately preceding when mortality is ascertained. The problem of attenuation bias due to noisy yearly data is tackled by computing long averages of yearly earnings.

Besides the contribution to the differential mortality literature, results from this paper can be used as an input in studies of progressivity of public programs such as Social Security and Medicare. Studies by Garret (1995), Gustman and Steinmeier (2001) and Armour and Pitts (2004) have analyzed how much of the progressivity built in the Social Security benefit formula remains once it is recognized that low-earners die faster than high earners. These researchers, in order to incorporate differential mortality, have used estimates of mortality differentials by income in the previous year from Kitagawa and Hauser (1973) and Duleep (1989). However, for reasons mentioned above, these estimates may not accurately represent differences in mortality rates across groups with different permanent income.[2]

It is not surprising that there are few studies that estimate differential mortality by some average of lagged earnings. To obtain these estimates very large micro datasets containing both earnings history and mortality status are required. The dataset most widely used in differential mortality studies in the United States (National Longitudinal Mortality Study) only reports cross-sectional income data. As an exception, Duleep (1986) used Social Security earnings data matched to mortality records to predict the death probability in a 5-year window (1973-1978) using a 5-year average of earnings

[2] Congressional Budget Office (2006) studied the effects of differential mortality on progressivity measures of the U.S. Social Security system using the Congressional Budget Office Long-Term model which incorporates in its microsimulation estimates of differential mortality by measures of lifetime earnings. Using the same model, Harris and Sabelhaus (2006) analyzed how changing the extent of differential mortality affects progressivity.

(1968-1972). Menchik (1993) used the National Longitudinal Survey of Mature Men and constructed a measured of average earnings up to age 61 to use as a control while probing for the effect of poverty on mortality. Finally, McDonough et al. (1997) employed data from the Panel Study of Income Dynamics to construct 10-year panels in which income averages are computed using the first 5 years and mortality status is ascertained over the subsequent 5 years.

In this paper, data from the 1984, 1993, 1996 and 2001 panels of the Survey of Income and Program Participation (SIPP) are matched to Social Security Administration (SSA) records on earnings, beneficiary, and mortality status. The resulting dataset contains roughly 138,000 individuals aged 35 to 84 for which the mortality window ranges from 3 to 20 years (depending on the SIPP panel), yielding a total of approximately 1.25 million person-year observations.

The breadth of the SIPP-SSA data set allows improvement upon previous studies in several different dimensions. First, more precise measures of lifetime earnings can be obtained as averages can use more than 5 years and also can be computed for ages that provide a better reflection of earnings potential. Second, I deal with the problem of reverse causation by not including observations of years immediately preceding the time window in which mortality will be ascertained. Third, while Duleep (1986) focused on white married men aged 35-65 and McDonough et al. (1997) pooled individuals aged 25 and older, in this paper separate models will be run by age and sex groups. Fourth, as the constructed data set encompasses 20 years of mortality status I am able to explore recent trends in the relationship between lifetime earnings and mortality. Fifth, given the sample size of the SIPP-SSA dataset, more precise estimates are obtained compared to those found in previous studies.

Though this data set constitutes a major innovation compared to previous ones, it has a limitation which is important to recognize. For the period 1951-1977, the earnings data corresponds to Social Security taxable earnings and then, it only includes earnings from Social Security covered jobs up to the taxable maximum.[3] Still, in Section II, I describe how the richness of the data set allows one to mitigate at least partly this limitation.

Findings regarding to the extent of differential mortality by lifetime earnings can be summarize as follows. First, there are large differentials in age-adjusted mortality rates across individuals in different quintiles of the own lifetime earnings distribution (e.g. men aged 35-49 in the bottom quintile have 6.2 larger age-adjusted mortality rates than those in the top quintile). Second, controlling for race, Hispanic origin, marital status and education only slightly reduces these differentials. Third, while differentials for men are (slightly) stronger by own compared to household lifetime earnings, the opposite is true for women.[4] Fourth, men and women have similar differentials when average household lifetime earnings are used to sort individuals in quintiles. Finally, differentials decrease very markedly with age.

With respect to trends in differential mortality by lifetime earnings, evidence is presented suggesting that it has increased in the 1984-2003 period. When sorting

[3] For the period 1978-2003 information on federal income taxable earnings is available, which includes earnings from jobs not covered by Social Security and it is not topcoded.
[4] Household earnings refers to the average lifetime earnings of the individual and his or her spouse (in case he or she is or was married).

individuals using own lifetime earnings, differentials increase for all age-sex groups except for men aged 35 to 49 and women 65 to 84. Moreover, if individuals are sorted by household lifetime earnings, I found that differentials have increased for all age groups except for men aged 35 to 49. This result should be interpreted with caution given that the quality of the earnings information in the data set increases over time and this could bias the results to finding increasing differential mortality. Still, given the far reaching implications of this finding on important and diverse issues such as future finances and progressivity of the Social Security system and how much of the income inequality is translated to differences in life expectancy across income groups, further work on this issue is definitely warranted.

II. Data

This study uses data from the 1984, 1993, 1996 and 2001 panels of the Survey of Income and Program Participation (SIPP) matched to several files administered by the Social Security Administration (SSA) containing information on earnings, disability and mortality.

The SIPP provides information for a representative sample of the U.S. non-institutional population. It contains information about cash and noncash income, taxes, assets, liabilities, demographics, labor force status and participation in government transfer programs. The survey is a continuous series of panels with sample size ranging from 14,000 to 36,700 households and was conducted annually between 1984 and 1993, and then in 1996 and 2001. Individuals in the SIPP panel are interviewed every 4 months for the duration of the panel (the surveys range in duration from 2.5 to 4 years depending on the panel).[5]

Four SSA files matched to the mentioned SIPP panels were also used and are described next. The Summary Earnings Record (SER) provides yearly Social Security taxable earnings for the period 1951-2003 as well as a variable which reports whether the individual paid Social Security taxes in each quarter (this information is useful to deal with the problem of topcoded earnings, as explained in Section II.2). The Detailed Earnings Record (DER) contains federal income taxable earnings for the period 1978-2003. The Master Beneficiary Record (MBR) provides information about Social Security benefits receipt. Finally, the Numident file, which is updated from the State and Territorial Bureaus of Vital Statistics, Veteran's Administration, SSA offices and other SSA administrative files, reports year of death.

II.1. Sample Construction

To create the sample for this study (the MORTALITY sample), I start constructing a panel data, in which the unit of observation is a person year, containing basic demographic and economic variables from the SIPP panels. For time-varying variables (education, marital status and spouse links), monthly information from the SIPP was used to construct yearly observations. Observations in which the person was 24 years or younger are dropped. Second, information on Social Security annual earnings from 1951-2003 (from the SER), federal income taxable earnings from 1978-2003 (DER), disability status (MBR) and year of death (Numident) is attached to the sample. Third, the

[5] For more details on the SIPP, see www.bls.census.gov/sipp/index.html.

resulting intermediate data set is “aged” completing missing years with information from the previous year up to year 2003 (or up to year of death if the person died before 2003).[6] While variables that are time-invariant (birth year, sex, race and Hispanic origin) are correct for the filled years, those that are time-variant (education, marital status and spouse links) could be wrong if there are changes in the individual situation. Lastly, only observations for individuals aged 35 to 84, born on 1907 or later are kept.[7]

Though this process of filling years with prior information introduces some measurement error in the education, marital and spouse links variable, the advantage is that it significantly enlarges the number of person-year observations in the data set. Also, note that the main variables in this study (mortality and measures of lifetime earnings) are unaffected by this decision. Finally, only 16 percent of the sample is aged more than 10 years (all individuals from the 1984 SIPP).

The resulting data set is a panel data where the unit of observation is a person-year and it includes yearly observations for individuals since the year they first entered the SIPP until 2003 (or until their death year, if they die before 2003).

II.2. Measures of Lifetime Earnings

This subsection describes how the measures of lifetime earnings used in the study are constructed. First, total annual earnings for years 1951-2003 are obtained. Second,

[6] That is, for an individual with SIPP data in 1984, 1985 and 1986, additional yearly observations for 1987 onwards are created using the variable values from 1986.
[7] Individuals younger than 35 are dropped because it is necessary to observe their earnings at ages while they were potentially attached to the labor market to construct the measures of lifetime earnings. Those older than 85 are eliminated from the sample to estimate more precisely age-specific mortality rates by groups (needed to construct mortality ratios). Finally, individuals born before 1907 are dropped because there is no earnings data for ages 44 and younger for them.

measures of lifetime earnings are constructed using 5 to 10 year averages of past indexed earnings. Lastly, quintiles of lifetime earnings by sex, 5-year age groups and 5-year calendar years are computed. Each of these steps is described next.

For the period 1951 to 1981 Social Security taxable earnings from the SER are used. A limitation of this earnings measure is that it is capped at the taxable maximum for each year. This problem is less severe for more recent years because the Social Security taxable maximum has been rising, in real terms, over time. However, for earlier years this problem is significant specially for men (for example, for men in the sample born between 1920 and 1924, 67 percent of them hit the taxable maximum in some year between 1960 and 1964). Fortunately, for years 1953 to 1977, the SER contains a variable called “Pattern of Quarters of Coverage” which reports whether an individual paid Social Security taxes for wages and salary in each quarter of the year. For individuals that have capped earnings, this variable can give bounds on their uncapped earnings assuming that their flow of earnings is constant across time.[8] Finally, total earnings for this study are set at the midpoint between these bounds (for example, for individuals hitting the taxable maximum in the second quarter, their annual earnings are set at 3 times the taxable maximum).[9] The procedure used to assign earnings above the taxable maximum, though its simplicity, constitutes a major improvement over other studies that have used this dataset given that exploratory analysis of the Patterns of

[8] For example, if an individual with a constant flow of earnings hit the taxable maximum in a year and he only paid Social Security taxes in the first two quarters, then we know that her uncapped earnings were at least twice but not more than four times the taxable maximum of that year (if the earnings were more than four times the taxable maximum, she would have hit it in the first quarter).

[9] For individuals that hit the taxable maximum in the first quarter, we cannot assign an upper bound. For these individuals I assume that their earnings were 8 times the taxable maximum. This assumption is somewhat innocuous given that for this study individuals are assigned to earnings quintiles of lifetime earnings.

Quarters of Coverage variable reveals that it contains significant information about individuals' uncapped earnings.[10]

For years 1981 to 2003 three annual earnings variables from the DER are used: IRS taxable income from wages and tips (box 1 of W-2 form); deferred wages (box 13); Medicare taxable self employment income (1040 SE). The sum of these earnings measures generates the value of total earnings used for this period. For years 1978-1980 information from the DER was not used because researchers familiar with this file believe that the quality of information on it was not very good for this period. Instead, for individuals who hit the taxable maximum in these years, the total earnings variable was set as the weighted average between total earnings in 1977 and 1981 (provided this amount was higher than the taxable maximum in these years).[11]

To construct the permanent earnings measure, an average of yearly past indexed earnings was computed. When constructing this measure the goal was to proxy the permanent earnings level of the individual while she has the closest attachment to the labor market. Also, in order to mitigate the problem of reverse causation, the measure does not include earnings received in the three years preceding when mortality is ascertained.

Taking these issues in consideration, for individuals aged 50 and older, the permanent earnings measure was constructed as the 10-year average earnings from ages 38 to 47. For individuals younger than 50, there is a trade-off between using more years

[10] On a sample of individuals aged 45 to 55 that hit the Social Security taxable maximum in 1984, regressing uncapped earnings in that year on sets of dummy variables reporting whether the individual hit the taxable maximum in the first quarter, second, third, fourth or not at all for years 1969 to 1977 yields an Adjusted R-square of about 0.3.

[11] The weights for year 1977 when imputing year 1978 was 3/4, for year 1979 was 2/4 and for 1980 1/4.

and reducing noise in the data and using fewer years to avoid including earnings at younger ages when high-earners may still be acquiring their educational degrees. This trade-off is resolved decreasing the number of years used in the computation of the average for younger ages (for those 35-45: 5 years are used to compute the average; for those 46-49: 6, 7, 8, 9 years, respectively).

Finally, as the goal is to estimate how mortality rates vary for individuals in different positions of the distribution of lifetime earnings, I sort individuals into quintiles of the lifetime earnings distribution by sex, 5-year age and 5-year cohorts. These quintiles are computed by own earnings and also by the average lifetime earnings between the individual's own lifetime earnings and his/her spouse (for simplicity, I call this measure the household lifetime earnings).

II.3. Is the Sample Representative?

The MORTALITY sample constructed for this study constitutes a unique data set to explore the relationship between lifetime earnings and mortality. However, the way that it was constructed (pooling SIPP panels, matching them to SSA records and "filling" years) may raise doubts with respect to the representativeness of sample. The question is whether the results are representative of certain period of time (i.e. the period of time when mortality was ascertained). As individuals in the MORTALITY data set enter the sample when they first were interviewed in the SIPP and remain in the sample until they die or the year 2003, the sample contains observations for years 1983 to 2003 but its composition is tilted towards later years.

To tackle this problem, and make the sample representative for the period 1983-2003, I obtained population counts by age, sex, race, Hispanic origin and year for the mentioned period from Census intercensal estimates.[12] The same age restriction used for constructing the MORTALITY sample was applied to the Census data (only individuals aged 35 to 84 were kept). Next, I constructed weights in order to match the distribution of observations in the sample by sex, 5-year age groups, race, Hispanic origin and 5-calendar year groups. Table 1 shows that the age, sex and race distribution in the unweighted MORTALITY sample is similar to the Census one. However, the distributions by year are very different (e.g., 43% of the observations in the MORTALITY sample correspond to the period 1998-2002, whereas in the Census data, 27% of the observations are in this group). Finally, comparing Columns 2 and 3, we can see that once the MORTALITY sample is reweighted, the distribution by age, sex, race, Hispanic origin and year matches closely the distribution in the Census data. Then, I used the constructed weights for all results presented in the remainder of the paper.

To assure that the sample is representative of the population for the 1983-2003 period, we can also compare, for particular years, the sample distribution by age, sex, race, Hispanic origin, education and marital status to the distribution from the SIPP. Using the SIPP as the benchmark allows us to compare a larger set of covariates. In this way we can check whether the two key steps in the construction of the data set (matching the SIPP pooled panel to SSA records and filling missing years) made the cross sectional patterns of the sample to diverge compared to those from a national representative sample. Table 2 presents this comparison for selected years 1984 and 1996 (for other

[12] The Census estimated counts were obtained at http://www.census.gov/popest/estimates.php.

years the same patterns emerge). The table shows that the sample closely replicates statistics from the SIPP in a particular point in time and also the changes in these distributions across time (for example, the fraction of the population with a college degree increased from 17.3% to 21.1% according to estimates from the SIPP and from 17.2% to 21.3% according to the sample).

Another important aspect to determine the reliability of the results from this study regards the quality of the mortality data. To gauge its quality, I compare sample death rates by age and sex to those computed using data from the Human Mortality Database (HMD) as a benchmark.[13] As mortality rates have decreased substantially for later cohorts, for each age-sex group, HMD death rates were constructed using rates by year of birth and computing the weighted average using as weights the cohort sample distribution. Figure 1 presents the sample and HMD mortality rates by age and sex for men and women. We can see that the sample mortality rates follow quite closely those from the HMD thought the sample rates seems slightly higher for ages 80-84.

III. The Extent of Differential Mortality by Lifetime Earnings

This section presents estimates of the extent of differential mortality by lifetime earnings. In the first subsection, mortality ratios are reported for groups defined by race, Hispanic origin, education, marital status, disability status and lifetime earnings quintiles. The ratios, computed separately for sex, represent the relationship between the mortality

[13] The Human Mortality Database is a joint project between the Department of Demography of the University of California, Berkeley and the Data Laboratory of the Max Planck Institute for Demographic Research. The constructed database contains original calculations of death rates and life tables for national populations. More information at www.mortality.org.

rates for each group (compared to the whole population) once the rates have been adjusted for differences in the age distribution between the particular group and the population. The second subsection focuses on differences in mortality rates by lifetime earnings using logistic regressions to adjust for different sets of covariates.

III.1. Mortality Ratios

The mortality ratio for a group in certain age group (e.g. black men aged 35-49) is computed in the following way:

$$Mortality\ Ratio_{BLACK\ MEN} = \frac{\sum_{a=35,\dots,49} weight_a * mortality\ rates\ black\ men_a}{\sum_{a=35,\dots,49} weight_a * mortality\ rates\ men_a}$$

where *mortality rates black men* $_a$ is the one-year age-specific mortality rate for black men aged a, *mortality rates men*$_a$ is the one-year age-specific mortality rate for all men aged a and *weight*$_a$ corresponds to the fraction of men aged a from all men in this age group in the sample.

Then, the numerator is the age-adjusted one-year mortality rate for black men while the denominator is just the average mortality rate for men in the sample. A ratio of 1 for certain group informs that, once we adjusted for differences in the age distribution, the group has the same mortality rate as all individuals in the sample of that sex. Finally, a ratio higher than one (e.g. 1.5) means that the group has a higher age-adjusted mortality rate than individuals of the same sex in the sample (50% higher).

Table 3 presents mortality ratios for men. The first column reports ratios for all men in the sample whereas Columns 2 to 4, reports for certain age groups (35-49, 50-64 and 65-84). Ratios by race, Hispanic origin, education, marital status and Social Security Disability Insurance status replicate the general patterns documented in previous studies on differential mortality. Focusing on individuals of all ages (Column 1), we see that blacks have a 36% higher age-adjusted mortality rate (compared to all men), Hispanics a 16% lower rate, while college graduates have a 34% lower mortality rate.[14] Being never married, separated/divorced or widowed is associated with a 40% higher mortality rate while individuals having ever receipted Social Security Disability Insurance (DI) have a 3.1 times larger mortality risk.[15]

Comparing Columns 2 to 4 of Table 3 we can see how the male mortality ratios vary across age groups. The results show that mortality differentials by race, education and marital status tend to dilute over time (i.e. mortality ratios converge towards 1 for older individuals). However, it seems that mortality differentials by education tend to persist for older individuals (e.g. college graduates aged 35-49 have 45% lower mortality rate and this figure falls only to 32% for those aged 65-84).

Similarly as with other covariates, the excess mortality rate associated with men ever on DI decreases as we focus on older individuals. However, individuals on DI aged 50-64 have an almost 13 times higher mortality rate (compared to men in that age group) whereas those aged 35-49 have an 8 times higher mortality rate.

[14] Along this subsection, for brevity, mortality rate refer to age-adjusted mortality rate.
[15] Mortality ratios for individuals currently on DI are not computed for age groups 35-84 and 65-84 because individuals on DI have their status updated to Social Security retirees when they turn 65.

The bottom panel of Table 3 presents mortality ratios by lifetime earnings quintiles computed by sex, 5-year age and 5-cohort groups. For quintiles computed by own or household lifetime earnings we observe similar patterns though the gradient is slightly stronger when using own lifetime earnings. Overall, there is a strong relationship between these measures of lifetime earnings and mortality. Individuals aged 35 to 49 in the bottom lifetime earnings quintile have a 130% higher mortality rate (compared to all men) while those in the top have a 63% lower. For a rough sense of the relative mortality predictive power of lifetime earnings, we see that high school dropouts aged 35 to 49 have a smaller excess mortality risk than individuals in the bottom quintile of the lifetime earnings distribution (similarly, college graduates have a smaller decrease in mortality risk compared to those in the top quintile).

The decrease in mortality differentials by age group is very strong. As the numerators of the mortality ratios for the different quintiles of lifetime earnings correspond to standardized age-adjusted mortality rates and the denominator correspond to just the average mortality rate in the sample (for the corresponding age group), then, we can compute the ratio of age-adjusted mortality rates between the bottom and top quintiles by just dividing the corresponding mortality ratios. While this ratio for men aged 35-49 is 6.2 (2.31/0.37), it drops significantly to 2.5 for men aged 50-64 and to only 1.2 for men aged 65-84.

Though a drop in the ratio is expected given that this pattern is also observed for other economic and demographic characteristics, still the fact that the drop is so large suggests that there can be other explanations beyond just an age effect. Given that the sample contains data for the period 1983 to 2003, individuals in the sample aged 35-49 were born between 1934 and 1968 while those aged 65-84 were born between 1906 and 1938. That means that when comparing mortality ratios across columns we are comparing individuals from different age groups but also cohorts. Section 4 aims to estimate whether there are cohort effects by exploiting the fact that the sample encompasses 20 years of mortality data.[16]

Table 4 presents mortality ratios for women. Overall, the patterns of differential mortality by race, Hispanic origin, education, marital and DI status found for men are also present for women except from certain differences which I describe next. First, Hispanic women have similar adjusted mortality rates compared to all women. Second, mortality rates differences across marital status are less pronounced especially for women aged 65 to 84. Third, the mortality "penalty" for being on DI or having ever been on DI is higher but still the patterns are similar.

To compare estimates of differential mortality by lifetime earnings between men and women, we can focus on the bottom panels of Tables 3 and 4. While the gradient is stronger for men (than women) when using own lifetime earnings, it is strikingly similar when using household lifetime earnings.[17] The former result should be expected given the higher attachment to the labor market for men (which suggests that men are relatively

[16] Another important difference when comparing across columns is that the measure of lifetime earnings for older groups correspond to earnings further back in the past. For example, for individuals aged 50, the measure of lifetime earnings was computed averaging earnings while the person was aged 38 to 47 and the same age range was used for all individuals older than 50. Then, the difference across columns could arise for a waning effect on mortality of earnings differentials measured at certain age.

[17] For example, the ratio of age-adjusted mortality rates for the bottom to top quintiles of own lifetime earnings is just 2.5 for women aged 35 to 49 compared to 6.2 for men in that age group.

better "sorted" when own lifetime earnings are used). However, the latter result is a novel finding which deserves further exploration in future work.

III.2. Logistic results

In this subsection I investigate whether the basic patterns about differential mortality by own lifetime earnings are affected by adjusting for different set of covariates. To do that, logistic models are run where the dependent variable is an indicator that equals one if the individual died in the next year and the key independent variable is the quintile of own lifetime earnings to which the individual is assigned. Odds ratios are estimated relative to individuals in the bottom quintile.

As expected, given that lifetime earnings quintiles are computed by sex, 5-year age and 5-year cohort, results from running models with no covariates are very similar to those when age and cohort are added linearly as controls. Moreover, they are also close to those obtained adding dummy variables for single year of age and cohort. Given this, Figure 2 presents odds ratios from specifications with just three sets of controls: a) age and cohort, b) age, cohort, race and marital status, c) age, cohort, race, marital status and education. Results with age, cohort and race are not presented because they are very similar to those when only age and cohort are added as controls.

The same basic patterns that were revealed when looking at mortality ratios also emerge in the logistic specifications: the relationship is stronger for men than for women and for young individuals than for older ones. However, there are certain refinements of these patterns that are noticeable in the graphs. For men aged 35 to 49 and 50 to 64, we observe that the risk of dying (relative to being in the bottom quintile) decreases monotonically as we focus on individuals in higher earnings quintiles. For men aged 65-84, the relative risk of dying for individuals in the second and third quintile is not statistically significantly different from one, but it is lower than one for those in higher quintiles. For women aged 35-49 and 50-64 we also observe a monotonic relationship between relative risk of dying and lifetime earnings quintiles, though for those aged 65-84, there is no statistically significant relationship between mortality and lifetime earnings quintiles.

Remarkably, all the uncovered patterns are quite robust to the addition of race, marital status and education controls. Figure 2 shows that while for men the degree of differential mortality by lifetime earnings seem to slightly decrease when we control for these factors, for women it slightly increases (when adding only race and marital status) or keeps virtually unchanged (when adding all the mentioned controls).

IV. Trends in Differential Mortality by Lifetime Earnings

Though there have been several studies aiming to estimate trends in differential mortality by education and last year income (see Preston and Elo, 1995; Feldman et al., 1989; Duleep, 1989) evidence on trends of differential mortality by lifetime earnings in the United States is lacking. In this section, I intend to fill this gap by comparing estimates of differential mortality by lifetime earnings of earlier cohorts to later cohorts.

For each of the six age-sex groups used along the study, I divided observations in two groups by their birth year. The cut-off birth year was selected to create two data sets roughly of the same size. The resulting datasets correspond to the EARLY and LATE cohorts. For example, for men aged 35-49, observations of individuals that were born between 1934 and 1954 correspond to the EARLY cohort while those that were born between 1955 and 1968 to the LATE cohort.

Figure 3 present logistic estimates of the one-year probability of dying on own lifetime earnings quintiles. In general, there is evidence of increasing differential mortality. Typically, the odd ratios for individuals in the second to top quintile (relative to those in the bottom quintile) are statistically significant lower for individuals in the LATE cohort compared to those in the EARLY cohort. This pattern is present in all age-sex groups except for men aged 35 to 49 and women aged 65 to 84, for which the estimates are not statistically significantly different.

Table 5 complements the results by presenting the odds ratio of the top quintile relative to the bottom for the six age-sex groups by own and by household lifetime earnings. Comparing the results by own lifetime earnings to those by household lifetime earnings, we see that the same general patterns with respect to trends emerge (though even for women aged 65 to 84 there is evidence of increasing differential mortality).

This evidence of rising differential mortality by lifetime earnings should be taken with caution. The quality of the earnings data used in the study is increasing with time and a decrease in classification error of individuals in earnings quintiles will produce increasing differential mortality across time even though the true correlation between mortality and lifetime earnings has not changed. Note that up to 1977 only Social Security earnings were used (which includes only Social Security covered earnings up to the taxable maximum) and after that federal taxable earnings were also available. Moreover, even for the period 1951-1977, the two problems with earnings data were becoming less important given that Social Security coverage rates went up through the period and the taxable maximum increased in real terms. As pointed in the data section, the problem of topcoded earnings was handled using additional information on the pattern of quarters of earnings. However, at this point, no attempt has been made to tackle the problem of increasing coverage rates and their implications on estimates on trends.

It is left for future work to better handle this issue. Three potential strategies can be followed. First, the rich information in the SIPP could be used to select a sub sample of individuals for which coverage rates has been more constant through time (e.g. dropping self-employed, government and state employees). Second, given that starting in 1984 there is very good information on earnings, the analysis could be re-run using a short average of past earnings that do not go further back than 1984. Third, using the information on the pattern of measurement error in the data, it may be possible to construct bounds on the estimates and in this way a more precise assessment of the trends in differential mortality can be presented.

V. Conclusions

This paper estimates the extent and trends of differential mortality by lifetime earnings using a very large panel data containing information on mortality, earnings history and demographic and economic characteristics. Measures of lifetime earnings are constructed in order to deal with the problems of reverse causation and noise in yearly earnings data present in estimates of differential mortality by last year income.

Summarizing the results, I found a strong relationship between mortality and lifetime earnings, robust to controlling for usual covariates, weaker for women than for men and decreasing with age. Also, evidence is presented suggesting that differential mortality by lifetime earnings has increased in the 1984-2003 period.

If the suggestive evidence presented in this study on increasing differential mortality is confirmed, a whole set of important questions could be addressed. In particular: what are the causes and consequences of increasing differential mortality by lifetime earnings? With respect to causes, the explanations that have been put forward to explain differential mortality can be used to check if they can explain the rise in this correlation. For example, a potential explanation for increasing differential mortality by lifetime earnings could be that the correlation between poor lifestyle habits (like smoking, poor diet and lack of exercise) and lifetime earnings also have increased over time. Another explanation could be that recent advances in medical treatments are more easily available to high earners than low earners relative to the past.

On the other hand, the implications of increasing differential mortality are far-reaching in different dimensions. First, if the “life-expectancy premium” for high-earners is increasing over time, this may worsen the expected deficits of the U.S. Social Security system given that high-earners receiving larger benefits will collect them (on average) for a longer period of time (Diamond and Orszag, 2004) . Second, studies that aimed to establish the progressivity of Social Security have used historical data on the correlation between earnings and mortality in order to account for the effect of differential mortality on progressivity measures. However, if differential mortality by lifetime earnings continues to increase over time, then we should expect that, given other factors constant, the progressivity of the system will worsen. Finally, for individuals concerned with the degree of income inequality in society, knowing that the correlation between earnings and mortality is increasing could be particularly troubling because it suggests that the inequality in current income is reflected in an increasing inequality in life expectancy across income groups.

References

Armour, Brian and M. Melinda Pitts (2004). “Incorporating Insurance Rates Estimates and Differential Mortality into the Net Marginal Social Security Tax Rate Calculation.” *Public Finance Review*, Vol. 32, No. 6 (November), pp. 588-609.

Congressional Budget Office (2006). *Is Social Security Progressive?* CBO Economic and Budget Issue Brief, Washington, D.C. (December).

Diamond, Peter and Peter Orszag (2004). *Saving Social Security: a Balanced Approach.* Washington, D.C.: Brookings Institution Press.

Duleep, Harriet Orcutt (1986). “Measuring the Effect of Income on Mortality using Longitudinal Administrative Record Data.” *Journal of Human Resources*, Vol. 21, No. 2 (Winter), pp. 238-51.

Duleep, Harriet Orcutt (1989). “Measuring Socioeconomic Differentials over Time.” *Demography*, Vol. 26, No. 2 (May), pp. 345-51.

Evans, William and Perry Singleton (2006). “A Reexamination of the Income and Mortality Gradient.” Manuscript, Department of Economics, University of Maryland.

Feinstein, Jonathan (1993). “The Relationship between Socioeconomic Status and Health: a Review of the Literature.” *The Milbank Quarterly*, Vol. 71, No. 2, pp. 279-322.

Feldman, Jacob, Diane Makuc, Joel Kleinman and Joan Cornory-Huntley (1989). “National Trends in Educational Differentials in Mortality.” *American Journal of Epidemiology*, Vol. 129, pp. 919-33.

Garrett, Daniel (1995). “The Effects of Differential Mortality Rates on the Progressivity of Social Security.” *Economic Enquiry*, Vol. 33 (July), pp. 457-75.

Gustman, Alan and Thomas Steinmeier (2001). “How Effective is Redistribution under the Social Security Benefit Formula?” *Journal of Public Economics*, Vol. 82, No. 1 (October), pp. 1-28.

Harris, Amy Rehder and John Sabelhaus (2006) “How Does Differential Mortality Affect Social Security Progressivity and Finances?” Technical Paper 2005-5. Congressional Budget Office, Washington, D.C. (March)

Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data downloaded on November, 2006).

Kitagawa, Evelyn and Philip Hauser (1973). *Differential Mortality in the United States: A Study in Socioeconomic Epidemiology*. Cambridge, MA: Harvard University Press.

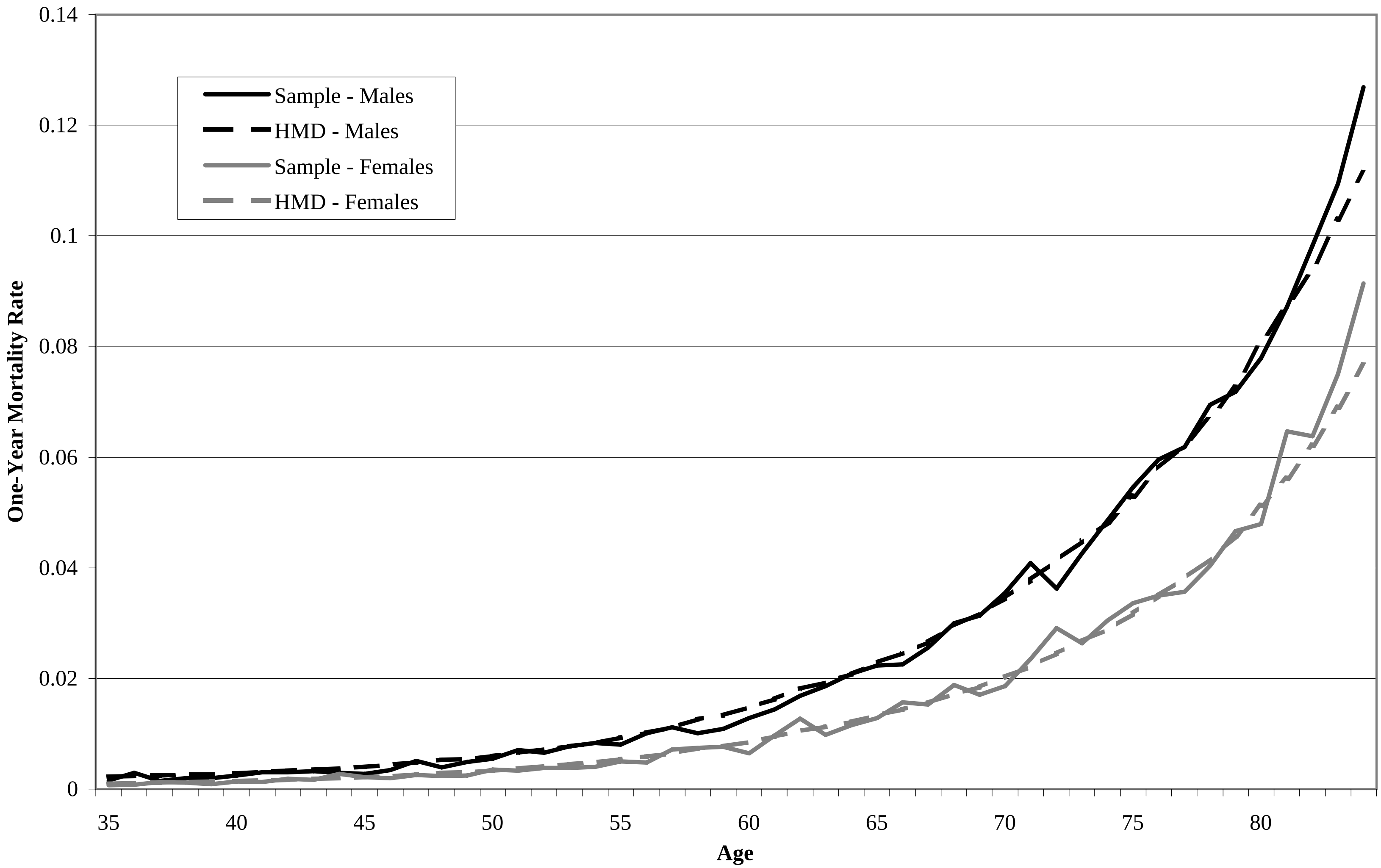
McDonough, Peggy, Greg Duncan, David Williams and James House (1997). “Income Dynamics and Adult Mortality in the United States, 1972 through 1989.” *American Journal of Public Health*, Vol. 87, No. 9 (September), pp. 1476-83.

Menchik, Paul (1993). “Economic Status as a Determinant of Mortality among Black and White Older Men: Does Poverty Kill?” *Population Studies*, Vol. 47, No. 3 (November), pp. 427-36.

Preston and Elo (1995). “Are Educational Differentials in Adult Mortality Increasing in the United States?" *Journal of Aging and Health*, Vol. 7, pp. 476-96.

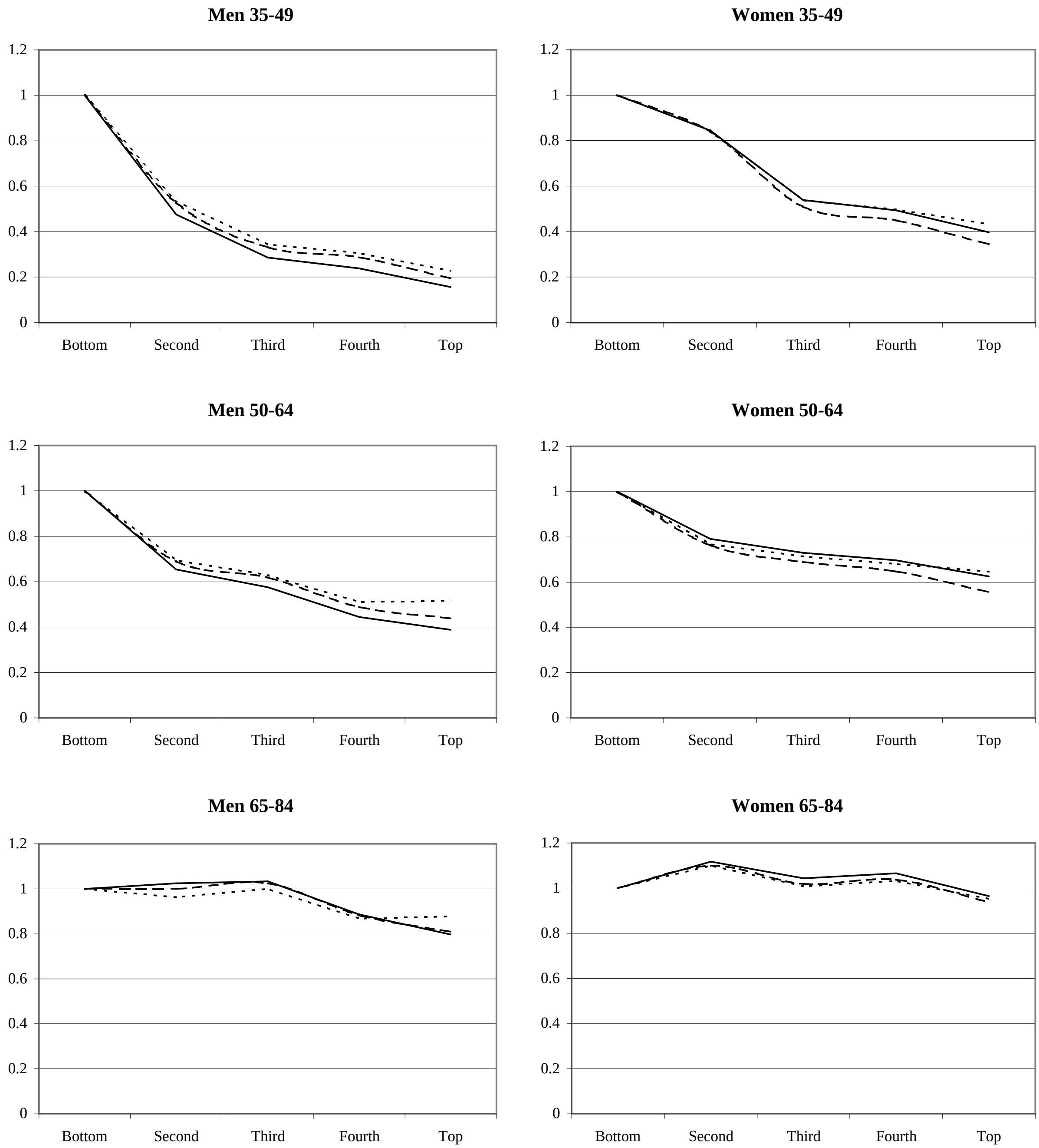
Sorlie, Paul, Eric Backlund and Jacob Keller (1995). “US Mortality by Economic, Demographic and Social Characteristics: The National Longitudinal Mortality Study.” *American Journal of Public Health*, Vol. 85, No. 7, pp. 949-56.

Figure 1. Sample Death Rates by Age and Sex. Comparison to Death Rates from HMD



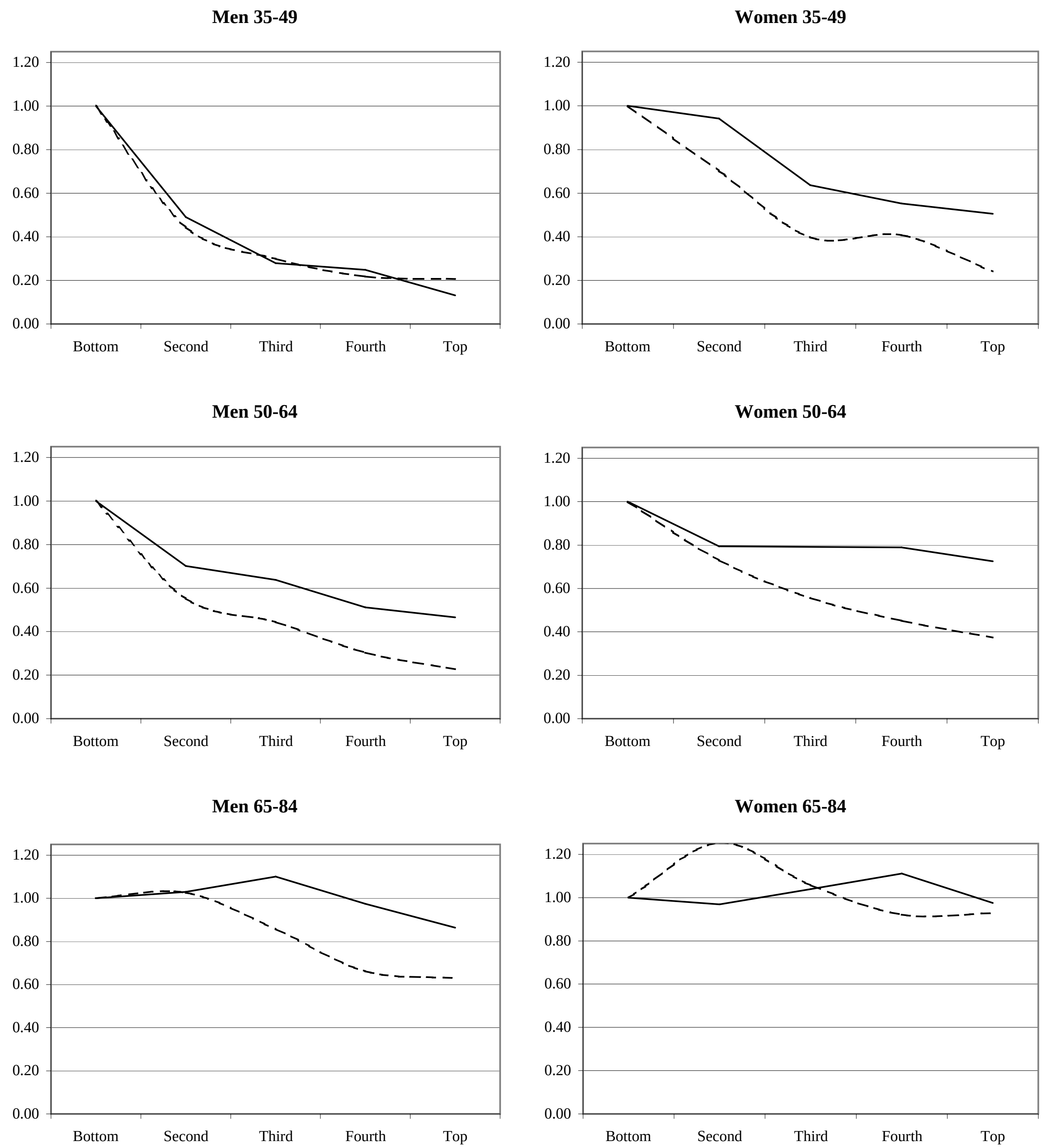
Note: HMD refers to the Human Mortality Database. For each age-sex group, HMD death rates were constructed using rates by year of birth and obtaining the weighted average using as weights the cohort sample distribution.

Figure 2. Adjusted Odds Ratios of One-Year Mortality by Own Lifetime Earnings Quintiles



Note: Adjusted odds ratios are obtained from logistic regressions of one-year mortality indicators on own lifetime earnings quintiles adjusting for age and birth year (solid line); age, birth year, race and marital status (dashed line); and age, birth year, race, marital status and education (dotted line).

Figure 3. Adjusted Odds Ratios of One-Year Mortality by Own Lifetime Earnings Quintiles
Late versus Early Cohorts



Note: For each age-sex group, observations where split to Early or Late Cohorts according to whether the person's birth year was above or below certain cut-off year. This cut-off year was selected to split the observations in two groups of roughly the same size. Adjusted odds ratios for the Early Cohort (solid line) and Late Cohort (dashed line) are obtained from logistic regressions of one-year mortality indicators on own lifetime earnings quintiles adjusting for age and birth year.

Table 1. Sample Descriptive Statistics. Comparison to Census Data

	MORTALITY Sample		Census
	Unweighted	Weighted	
% Male	46.8	47.1	47.2
Average age	53.6	53.3	53.2
% Age 35-49	45.6	46.4	46.5
% Age 50-64	31.4	30.4	30.5
% Age 65-84	23.0	23.2	23.0
Race Groups			
% White	86.6	86.1	86.0
% Black	10.0	10.2	10.3
% Other race	3.4	3.7	3.7
% Hispanic	6.5	7.3	7.2
Observation year			
Average	1996.6	1993.7	1993.8
% Year 1983-1987	8.8	19.9	19.8
% Year 1988-1992	11.4	22.4	22.4
% Year 1993-1997	27.5	24.8	24.8
% Year 1998-2002	42.7	27.2	27.3
% Year 2003	9.5	5.7	5.7

Note: Census male, age and race statistics correspond to average yearly statistics weighted by population counts for each year in the period 1983-2003. The weights used in the MORTALITY Sample Weighted column were constructed in order to match the sample distribution by sex, 5-year age groups, race, Hispanic origin and 5-calendar year to the Census counts in the period 1983-2003.

Table 2. Sample Descriptive Statistics. Years 1984 and 1996. Comparison to SIPP Data

	Year 1984		Year 1996	
	MORTALITY Sample	SIPP	MORTALITY Sample	SIPP
% Male	46.8	46.5	47.2	45.7
Average age	53.3	53.0	53.1	53.7
% Age 35-49	43.3	43.9	48.5	47.8
% Age 50-64	34.0	34.6	28.3	27.3
% Age 65-84	22.7	21.5	23.2	24.9
Race Groups				
% White	87.8	88.2	85.7	84.0
% Black	9.7	9.5	10.4	11.8
% Other race	2.5	2.4	3.9	4.2
% Hispanic	5.4	4.5	7.6	7.8
Education				
% Less than High School	31.9	32.1	21.8	21.1
% High School	34.1	34.1	33.0	32.0
% Some college	16.7	16.5	23.9	25.9
% College	17.2	17.3	21.3	21.1
Marital Status				
% Never married	5.4	5.3	9.6	8.1
% Married	73.0	73.1	66.6	64.6
% Separated/Divorced	11.8	12.0	14.7	15.7
% Widowed	9.9	9.6	9.2	11.5

Note: MORTALITY Sample statistics where computed using weights in order to match the sample distribution by sex, 5-year age groups, race, Hispanic origin and 5-calendar years to the Census counts in the period 1983-2003.

Table 3. Mortality Ratios - Men

Age group		35-84	35-49	50-64	65-84
All		1.00	1.00	1.00	1.00
Race Groups					
White		0.97	0.90	0.95	0.99
Black		1.36	1.74	1.58	1.24
Other race		0.72	1.12	0.77	0.65
Hispanic		0.84	0.98	0.94	0.79
Education					
Less than High School		1.27	1.56	1.35	1.20
High School		1.00	1.11	1.05	0.97
Some college		0.91	0.97	0.89	0.90
College		0.66	0.55	0.64	0.68
Marital Status					
Never married		1.37	1.96	1.64	1.19
Married		0.88	0.72	0.85	0.91
Separated/Divorced		1.38	1.56	1.46	1.33
Widowed		1.44	1.54	1.95	1.26
Disability Insurance					
Currently on DI		-	8.22	12.90	-
Ever on DI		-	8.23	4.18	-
Lifetime Earnings Quintiles					
Bottom	Own	1.31	2.31	1.62	1.05
	Household	1.24	2.15	1.52	1.02
Second	Own	1.08	1.11	1.07	1.08
	Household	1.04	1.16	1.04	1.03
Third	Own	1.01	0.67	0.94	1.08
	Household	0.94	0.79	0.83	1.00
Fourth	Own	0.85	0.56	0.73	0.94
	Household	0.92	0.53	0.82	1.01
Top	Own	0.75	0.37	0.64	0.85
	Household	0.85	0.38	0.78	0.93

Note: The mortality ratio for a group is computed by dividing the weigthed average of the one-year age-specific mortality rate for the group, where the weights correspond to the fraction of men in the sample in that age, by the male mortality rate in the sample. DI corresponds to Social Security Disability Insurance.

Table 4. Mortality Ratios - Women

Age group		All: 35-84	35-49	50-64	65-84
All		1.00	1.00	1.00	1.00
Race Groups					
White		0.97	0.93	0.93	0.98
Black		1.31	1.53	1.58	1.22
Other race		0.95	0.86	0.98	0.95
Hispanic		0.99	0.92	0.97	1.00
Education					
Less than High School		1.26	1.60	1.49	1.16
High School		0.94	1.12	0.89	0.94
Some college		0.86	0.78	0.82	0.88
College		0.73	0.58	0.64	0.77
Marital Status					
Never married		1.22	1.91	1.59	1.06
Married		0.85	0.75	0.81	0.86
Separated/Divorced		1.20	1.34	1.32	1.16
Widowed		1.21	1.83	1.43	1.10
Disability Insurance					
Currently on DI		-	10.27	16.26	-
Ever on DI		-	10.98	4.54	-
Lifetime Earnings Quintiles					
Bottom	Own	1.09	1.53	1.30	0.99
	Household	1.21	1.95	1.46	1.08
Second	Own	1.04	1.29	1.01	1.03
	Household	1.05	1.11	1.14	1.03
Third	Own	0.99	0.82	0.94	1.02
	Household	0.98	0.77	0.89	1.02
Fourth	Own	0.99	0.76	0.90	1.04
	Household	0.90	0.71	0.73	0.97
Top	Own	0.90	0.60	0.81	0.95
	Household	0.84	0.48	0.78	0.89

Note: The mortality ratio for a group is computed by dividing the weigthed average of the one-year age-specific mortality rate for the group, where the weights correspond to the fraction of women in the sample in that age, by the female mortality rate in the sample. DI corresponds to Social Security Disability Insurance.

Table 5. Trends in Differential Mortality by Lifetime Earnings
Estimated Odds Ratios of One-Year Mortality. Top relative to Bottom Quintile

		Own		Household	
		Early Cohort	Late Cohort	Early Cohort	Late Cohort
Men	35-49	0.13	0.21	0.14	0.25
	50-64	0.47	0.23 **	0.63	0.28 **
	65-84	0.86	0.63 **	0.98	0.74 **
Women	35-49	0.51	0.24 *	0.29	0.17
	50-64	0.73	0.37 **	0.69	0.21 **
	65-84	0.98	0.93	0.90	0.62 **

*: Significantly different to the Early Cohort estimates at the 5% level.
**: Significantly different to the Early Cohort estimates at the 1% level.
Note: For each age-sex group, observations where split to Early or Late Cohorts according to whether the person's birth year was above or below certain cut-off year. This cut-off year was selected to split the observations in two groups of roughly the same size.