Estimating the effect of income on health and mortality using lottery prizes as exogenous source of variation in income

Lindahl, M.

Estimating the Effect of Income on Health and Mortality using Lottery Prizes as Exogenous Source of Variation in Income

ABSTRACT

A vast literature has established a strong positive relation between income and health status and a negative relation with mortality. This paper studies the effects of income on health and mortality, using only the part of income variation due to a truly exogenous factor: monetary lottery prizes of individuals. The findings are that higher income causally generates good health and that this effect is of a similar magnitude as when traditional estimation techniques are used. A 10-percent income increase improves health by about 4-5 percent of a standard deviation.

1. Introduction

It is well known that individuals with high socioeconomic status have better health.[1] This appears to be true for most measures of socioeconomic status, such as education, income and occupation, in most regions or countries and for most measures of health and mortality. Whether these associations can be interpreted as causal effects of the socioeconomic status measures on health is more dubious, however. As formulated by Deaton & Paxson (1998): "There is a well-documented but poorly understood 'gradient' linking socioeconomic status to a wide range of health outcomes". Distinguishing an association from a causal relation is vital for policy purposes. If income causally determines health, an evaluation of any policy affecting people's income should take into account its effect on their health.

This paper focuses on whether disposable income has a causal effect on health.[2] There are three main reasons why a positive relation between income and good health is observed. First, this might be due to a spurious association driven by factors such as genetic or social background which are likely to affect both income and health. Second, it might be due to reverse causation, for example through bad health decreasing work productivity or hours of work, which would reduce income.[3] Third, it might be due to a causal effect of income on health. If such an effect is found, the next step is to investigate through which channels it operates.[4] If no causal effect is found, policy makers can instead focus on policies with a direct effect on health (such as health care reforms) or an indirect effect through other factors than income.

Simply estimating health variables as a function of income is likely to be insufficient for consistently estimating the causal effect of income on health. The reverse causality issue and/or insufficient control variables create a need for alternative identification strategies. This study estimates the causal effect of income on health using an identification strategy not previously applied to this issue. I use information on monetary lottery prizes to create exogenous variation in income, an approach suggested by Smith (1999). Using lottery prizes is very appealing since, by definition, a lottery randomly draws winners from a pool of participants. If all participants have the same chance of winning (which is true if they all buy the same number of equally priced lottery tickets and the drawing is correctly administrated), lottery prizes create exogenous variation in income among individuals.

I use data from the Swedish Level of Living Surveys (SLLS) 1968, 1974 and 1981. SLLS contains of a vast number of health questions, and register information on the death date of individuals. As health measures in 1981, I use a standardized index of bad health, constructed from the questions on health symptoms, indexes of separate health symptoms and measures of morbidity. I am able to calculate the amount won on lotteries by the respondents during a 13-year period (1969-1981). Register information on income has then been matched against the information in SLLS. I calculate respondents' disposable family income spanning over 15 years (1967-1981).

The next section discusses the data and some conceptual issues. Section 3 investigates the exogenous nature of player status and monetary lottery prizes. In section 4, health and mortality equations are estimated as a function of lottery prizes. I also compare estimated effects of income on health using traditional techniques to utilizing monetary lottery prizes as exogenous variation in income. The last section draws conclusions and discusses the findings.

2. Data, variable construction and some conceptual issues

The data set is constructed from the Swedish Level of Living Survey (SLLS) data base for 1968, 1974, and 1981.[5] SLLS follows individuals across waves, so that many individuals are included in all years. New individuals are added in each wave to maintain a representative sample.[6] An advantage of using the SLLS database is that it also contains extensive questions on health as well as socioeconomic status variables, and that it has been matched with register data for income and the respondents' death status/date. All three waves of SLLS also contain a question on the amount of money won on lotteries.

Health/Mortality

All waves of the SLLS data set contain a large number of questions regarding health symptoms, for example direct questions on sicknesses (ranging from coughs to cancer) as well as questions on other health-related conditions (for instance, limitations in the ability to move and pain in the back). To simplify and condense the presentation, I attempt to simultaneously capture all aspects of health status by combining 48 health symptoms to an interpretable overall measure of the general health status. Based on these symptoms, I construct a Standardized index of bad health (STDH), which I then use as the dependent variable in the analysis of the effect of income on health. For a description of how STDH is constructed, see Appendix 1, where I am also discussing the pros and cons of using this measure.

Descriptive statistics for the number of bad health symptoms and STDH are shown in Table 1. The mean individual has almost 6 symptoms, and the distribution is left- skewed. STDH is standardized, with the mean zero and the standard deviation one, but somewhat left-skewed. For the mean individual, a one-tenth standard deviations increase in bad health is equivalent to a move from the 63rd to the 66th percentile, which is comparable to an increase in the number of health symptoms from 6 to 7. I also use separate/grouped health variables to capture specific symptoms/diseases. These are the number of symptoms for immobility (coded from 0 to 3), poor mental health (0 to 5) and cardiovascular diseases (0 to 7).[7] I also use a measure of no, some or severe headache and overweight, respectively. Table 1 shows the average number of symptoms for immobility and poor mental health to be around 0.5, whereas it is somewhat higher for number of cardiovascular symptoms. These three variables are count variables and I therefore use Poisson as well as OLS regressions. Both overweight and headache are coded as 0 to 2, with averages about 0.2 and 0.5, respectively. Since these are ordinal variables, I estimate ordered Probit as well as OLS models.

The SLLS data has also been matched with register data on death dates for individuals deceased before January 1 1997. I construct two mortality indicators to be used as dependent variables: D5 (and D10), which equal one if the individual died within 5 (10) years after the last interview date in 1981 and zero otherwise. Note that STDH is based on the individuals' subjective responses to questions on health symptoms. I do not have access to any objective measure of health symptoms in this data. However, the mortality measures are based on objective (i.e. register) information and will also capture some additional aspects of health, i.e. those leading to death.

In Table 2, I correlate some health (measured in 1981) and mortality measures. As can be seen, STDH is highly correlated with the number of health-related symptoms (SYM) as well as the specific group of symptoms in 1981. STDH is more strongly correlated with the mortality measure and the number of visits to the doctor than what is the case for SYM. This indicates that in addition to reflecting the large number of different health symptoms, the weighting scheme also captures other features related to general health.

Lottery Prize

In the SLLS for 1968, 1974, and 1981, the respondents were asked: "have you ever in your life won at least 1,000 Swedish Kronor (SEK) on gambling or lottery of any kind?".[8] If the answer to this question was 'yes', the respondents were also asked: "approximately how much altogether?" The answer to the last question is here interpreted as a statement of the sum of all monetary lottery prizes won by the respondent has won until the time of the survey. This information is then used to construct the lottery prize measures.[9]

Using the information from these lottery questions entails a couple of potential weaknesses. First, the second question is somewhat unclear. How much the respondent has won altogether might mean the total sum at the time of the largest win. Interpreting the question in this way means that the lottery prize measure should have been constructed somewhat differently. Using such an alternative lottery measure produces very similar results, however.[10]

Second, the SLLS did not contain a question of how often or how much the respondent plays on lotteries. Potentially, this is a great disadvantage, since people playing on lotteries are likely to play different amounts. If the respondent considered the question to concern the sum of all lottery prizes won, and did not subtract the money played for, the lottery variable would be expressed in gross, instead of net, terms. Section 3 contains a discussion of these issues.

Third, no question was included on when the respondent won on the lotteries. In order to create some time limit for when prizes were won, I use the difference between prizes stated by the respondent between two consecutive SLLS surveys as an estimate of how much the respondent has won since the previous survey.[11] The lottery prize 1969-1981 is calculated as $\overline{L}_{81,13} = \overline{L}_{74,6} + \overline{L}_{81,7}$. $\overline{L}_{74,6}$ is the average lottery prize 1969-1974, calculated as the sum of lottery prizes between SLLS 1968 and 1974 divided by the number of years since the previous survey (i.e. 6 years) and $\overline{L}_{81,7}$ is the average lottery prize 1975-1981, calculated as the sum of lottery prizes between SLLS 1974 and 1981 divided by the number of years since the previous survey (i.e. 7 years). In order to get a lottery prize in 1998 prices, I adjust the prizes using Statistics Sweden CPI figures for the midpoint between these years (July 1971, and December 1978) as base years.[12]

Fourth, the SLLS did not contain a direct question about whether the respondent plays on lotteries, that is, I do not know which individuals participated in the "lottery experiment". I can, however, isolate the individuals known to have played the lotteries since I know who claimed to have ever won at least 1000 SEK. These guaranteed lottery players are then contrasted against the inseparable group of non-players and those who played but never won. In the following, the first group is labeled as players, and the second as non-players.

As shown in Table 1, 21 percent of the full sample are players. 113 of the 626 individuals who won lottery prizes 1969-1981, claimed to also have won on lotteries in SLLS in 1968. Hence, these individuals are counted to also have been players before 1969. Since having won prizes both before and after 1969 probably indicates that an individual is playing a great deal, I include an indicator for these individuals in the regressions.

Are these potential weaknesses in the lottery information likely to seriously affect the results? Quite strong evidence for this to be unlikely is found in Imbens et al. (2001), which analyzes the effect of unearned income on labor earnings, savings and consumption, using information on monetary lottery prize winners in the US. They have information on which year a prize was won and the number of tickets bought. In their data, small-prize winners buy fewer tickets than medium- and big-prize winners, but there is no significant difference between the last two groups. They also find that the number of tickets bought is not significantly correlated with earnings, and that the estimates from regressions of earnings on lottery prizes are very similar if controls such as the number of tickets bought and the year of winning are included as controls.

Among the players, the average yearly lottery prize is 2000 SEK between 1969 and 1981. Note that 72 individuals won positive lottery prizes in both periods (1969-1974 and 1975-1981). A comparison of the characteristics of lottery and non-lottery players will be made in section 3. There, I will also show more statistics for lottery players, conduct a detailed analysis of what determines the player status and the amount won and also provide some aggregate evidence on the amount and type of lottery playing in Sweden in this period.

Income

The SLLS also has detailed information from tax registers which has been matched against the individuals by Statistics Sweden. This includes income from several different sources, such as work, capital and government transfers, and information on the amount of taxes paid (see Björklund & Palme, 2002, for details). This information is available from the tax registers from 1974 and onwards. These income components are also available for 1967 and 1973 from the SLLS-surveys conducted in 1968 and 1974. Hence, I have comparable measures for these income components for 1967 and 1973-81, mainly based on registers. Using this, and the lottery prize information for the periods 1969-1974 and 1975-1981, I can construct disposable family income measures for basically all individuals in the sample.

Since I do not know the lottery prize for each year (but only the amount won for the periods between the survey years), I calculate the average disposable family income (in adult equivalents) between year t-k and t, as $\bar{I}_{t,k} = \sum_{j=t-k+1}^{t} (y_j + \bar{L}_{t,k}) / a_j \Big/ k$. Family net income in year t, y_j, is calculated as tax-assessed income minus taxes plus transfers to the family, where the tax-assessed income includes pensions, sick pay, and unemployment benefits. Transfers include child and housing allowances; $\bar{L}_{t,k}$ is the individual's monetary lottery prize from year t-k to year t, divided by k. Both the family net income and the monetary lottery prize are already adjusted for inflation; a_j equals the square root of the number of household members;[13] t=1974,1981 and k=6, 7. The most permanent income measure available is the average income 1967-1981, which is used in the main analysis and calculated as $\bar{I}_{81,15} = \sum_{j=67}^{81} (y_j + \bar{L}_{74,6} + \bar{L}_{81,7}) / a_j \Big/ 10$. Since I only have information for 1967 and 1973-1981, I sum over these years and divide by 10.[14] All income measures are expressed in 1998 SEK prices.

Descriptive statistics for the main income variables are shown in Table 1, with the absolute values (in 10,000 SEK) shown in brackets. The mean and standard deviation are lowest for the most permanent measure. The standard deviation for this variable is very low, which can be illustrated by noting that doubling the income for the median individual brings him/her into the 99th percentile. Comparable yearly income measures for 1967, 1973 and

1974 are listed at the bottom of Table 1. Note that these income measures do not include lottery prizes, however. Income has increased from 1967 to 1981, which is partly due to people in the sample getting older, and partly to a real increase in disposable income during these years.

Additional variables

The other variables shown in Table 1 are used as controls in some of the later estimations. First, there is a group of variables including age and variables constant for each individual over time. The average person in the sample is 53 years of age in 1981 and the number of women is slightly higher than the number of men. The share of foreigners, defined as people having immigrated at any age, is about 5 percent. 27 percent grew up in families with difficult economic conditions and 21 percent grew up in a family where they themselves, any sibling or any parent had a serious or long lasting sickness.

Second, there are some variables capturing five socioeconomic characteristics of the respondent at the date of the earlier surveys in 1968 and 1974. On average, the individuals had 8.5 years of schooling, 74 percent were working and 79 percent were married or cohabiting in 1968. Six years later, these numbers remained quite similar, although average years of schooling increased somewhat. In order to proxy for the wealth of the individuals, I use a question where the respondents were asked to report whether they would have difficulties in bringing forth about SEK 12,000, in SEK 1998 prices, within one week.[15] The responses are divided into three groups. Those individuals who were unable to bring forth this amount are coded as being poor and thereby having "Very low wealth." Those individuals who could raise this amount of money by loans or in some other way (not including drafts from their own bank account) are coded as having “Low wealth.” The reference group consists of individuals who could bring forth this amount of money themselves. As can be seen from Table 1, the share of people with very low or low wealth decreased from 44 to 31 percent between 1968 and 1974.

Third, there are variables capturing lagged health and income, which are used as controls in some of the estimations. I note that inequality in health, measured by the standard deviation of number of sickness symptoms has increased over the years, whereas the inequality in disposable family income has fluctuated for this sample. The average individual has more sickness symptoms and a higher disposable family income in later years.

3. Players and lottery winnings

In this section, I look at some aggregate numbers on lottery playing in Sweden, the difference in characteristics between players and non-players and how player status is determined. I also look at what determines the amount of lottery prize won for the sample of players.

3.1. Evidence on the amount and types of playing in Sweden

To understand what is captured in the respondents’ stated lottery winnings used in this paper, Table 3 shows some aggregate statistics on the amount and types of playing in Sweden 1969-1981. I have divided the amounts into gambling and lotteries. The first category comes from betting on horse races and soccer, the only existing money gambles that was allowed in Sweden at that time. The individuals winning on gambling are not randomly drawn and hence, winnings could be related to some skill. The second category is from playing on number lotteries where each lottery ticket gives an identical probability of winning a prize.[16] As can be seen, roughly the same amounts are played on gambling and lotteries. For gambling, the amount actually paid back to the winners (after taxes and profit) is about 35 percent and for lotteries almost 50 percent of the amounts in Table 3. Hence, the amounts claimed to have been won in SLLS 1968, 1974 and 1981, are likely to reflect both winnings from gambling and lotteries but mainly winnings from lotteries.

Another issue is whether the aggregate figures are comparable to the amounts won in the sample in SLLS? I first note that my sample of 2,948 individuals is a sample of 0.07 percent of the population in Sweden in 1975.[17] In SLLS, the yearly average amount won is 2,500 SEK (in 1998 prices) for the sample of players (see Table 1). Hence, the sum of lottery prizes won in 1975, the mid-year of 1969-1981, is about 1,565,000 SEK for the full sample.[18] Multiplying the sum of lottery winnings in SLLS for the full sample by $100 \cdot (1/0.07)$ gives 2,280 MSEK (1,000,000 SEK=1 MSEK), which is the expected amount won on lotteries for the entire population, based on the sample in this study. The total amount paid out to winners in the population in 1975 was about 590 MSEK,[19] which amounts to 2,500 MSEK in 1998 prices. This number is very similar to the amount of winnings in SLLS (2,280 MSEK).[20]

3.2. Determinants of player status

Table 1 also shows descriptive statistics separated for players and non-players, where the last column contains p-values from a test of mean equality between the two groups. We see that players have, on average, a significantly higher income. This is also true for income in 1973 and 1974, which does not contain lottery winnings. Most health measures do not differ significantly between players and non-players.[21] Players are also more likely to be male, as well as, being single, working and with very low wealth in 1968.

Table 4 presents estimates from regressing a dummy for player status (=1 if the individual is a player) on a number of covariates, using a linear-probability model. In column 1, the probability of playing on lotteries first increases and then decreases with age and is much higher for men. Adding family background, pre-determined socioeconomic variables, STDH in 1968 and log income in 1967 gives that the probability of playing increases with being single and having bad health in 1968. There is also some evidence of the probability of playing being higher for individuals with health problems when growing up and lower for those with economic problems when growing up. Both lagged income and the proxies for wealth are insignificantly related to the probability of playing. The p-values from partial F-tests show the coefficients to be jointly significantly different from zero, even if women and the indicator for playing prior to 1969 are included in the regression, but excluded in the test.

Since some observable characteristics are related to the probability of playing on lotteries, the sample is restricted to players in the following analysis, which reduces the sample to just over one fifth of the original. However, there is also an intuitive reason for only including lottery winners. Since this study aims at mimicking an experiment, where money is randomly given to individuals, I like to include only individuals participating in the experiment. Otherwise, I would implicitly assume that individuals who participated (i.e. played but did not win), would have the same characteristics as those who choosing not to participate in the experiment. Notably, the results from the estimations using the lottery prize variable in section 4 remain basically unchanged if the whole sample is used.

3.3. Is the amount of monetary lottery prize won really exogenous?

Returning to Table 1, we see that for the sample of players, the average yearly lottery prize amounts to about SEK 2,500 (about $250 in 1998 value) per year. This figure is probably comparable to a policy change in taxes and transfers of quite realistic magnitude. The number corresponds to an average total amount of lottery prizes of SEK 32,500 in 1969-1981. During this period 305 individuals won positive lottery prizes amounting to less than SEK 10,000 and 38 individuals won a lottery prize amounting to more than SEK 100,000. These figures are in 1998 prices.

In Table 5, monetary lottery prizes for players are regressed on the same covariates as in Table 4. In column 1, the monetary lottery prize during the whole period is the dependent variable. The lottery prize is significantly higher if the player is a man, had no health problem in the family when growing up, did not work (marginally significant), did not have a very low economic status at the beginning of the period and has previously won lottery prizes. R^2 is not very high (0.020), even with this full set of variables. Note that we cannot reject that all variables, excluding women from the test, have no joint effect on lottery prizes.[22]

Columns 2-5 of Table 5 present regression results separately by periods. Columns 2-3 show results for lottery prizes won in 1969-1974 and columns 4-5 for 1975-1981. OLS are used in columns 2 and 4 whereas in columns 3 and 5 Tobit models are estimated. The reason for using a Tobit here is that although all individuals in these regressions are guaranteed to be players during 1969-1981, some individuals have not won any amount in one of these periods (1969-1974 or 1975-1981). Those individuals who won no prize in any of these periods, but still played, might have lost relatively more money. In that case, the lottery prize variable is censored just below the lowest positive value. The results from columns 2-5 are that income is related to lottery prizes in the first, but not in the second period (the p-value of a joint test for income 1967, 1973-1974 in columns 5 and 6 are 0.22 and 0.52 respectively).

In columns 5-6 I add a variable for the amount of monetary lottery prize won in 1969-1974. The estimate reveals that previous lottery prizes are unrelated to the lottery prize 1975-1981. This is a comforting result, since if lottery prizes reflect betting or if some individuals play more often than others or for more money (i.e. buy more lottery tickets), we would expect previous amounts of lottery prizes won to be positively correlated with later ones. Using a Tobit, the lagged lottery prize is negatively, but still insignificantly, related to lottery prizes won in 1975-1981. Note that both lottery prizes won 1969-1974 and being a player before 1969 are jointly insignificant (p-values 0.58 and 0.37 in columns 5 and 6).

Altogether, the results in this section support the usage of lottery prizes in analyzing the effect of income on health. First, health at the beginning of the period has no significant effect on the size of monetary lottery prizes. This is an important result, since it means that monetary lottery prizes appear to be exogenous with respect to health status. Second, there is no evidence of previous disposable family income being associated with the amount of lottery prize won, which suggests that high income people do not buy more lottery tickets, thereby increasing their chance of winning. This is also supported by years of schooling being insignificant, although this is somewhat contradicted by the economic status indicators in the regressions in Table 5. Third, previous lottery winnings are not significantly related to current lottery winnings. I therefore conclude that there is surprisingly little evidence in favor of lottery prizes being non-randomly distributed among individuals who play on lotteries, at least if controlling for the individuals' gender. Hence, the absence of information on the amount played seems to be a minor problem.

4. Estimating causal effects of income on health and mortality

4.1. Health and mortality regressions

I first estimate regressions of health in 1981 on the average lottery prize 1969-1981 and other covariates. The basic equation to be estimated is:

(1) $$H_{i81} = a + b\,\overline{L}_{i81,13} + q'X_{it} + h_{i81},$$

where H_{i81} is the standardized index of bad health (STDH) for individual i in 1981; $\overline{L}_{i81,13}$ is the average lottery prize 1969-1981 (see section 2 for an exact definition) in 10,000 SEK; X_{it} is a vector of demographic (a cubic in age and an indicator for women) and family background (indicators for being foreign, having had a bad economic situation and bad health in the family when growing up) variables, as well as, socioeconomic variables measured in 1968 (years of schooling, indicators for worker and marital status, low wealth and very low wealth), the previous health status in 1968 and an indicator for winning on lotteries in 1968 or earlier and; h_{i81} is a random error term.[23] The standard errors are robust to heteroscedasticity.[24] Note that the reason why I control for variables measured as early as 1968 and not later, is that variables measured later are potentially endogenous with respect to lottery prizes before 1969-1981. Also note that in the tables, a negative sign indicates a positive effect of income/lottery prize on health, since the dependent variable is in bad health.

The first three columns of row 1 of Table 6 show estimates, based on the estimation of equation 1 for all players, with less and more covariates. With only women and the cubic in age as controls, winning 10,000 SEK per year 1969-1981 is estimated to increase health by 0.07 standard deviations (column 1) and this estimate is significant. Adding family background, socioeconomic variables as well as health at the beginning of the period reduces the estimate to 0.05 (column 2). Adding an indicator for winning on lottery before 1969 barely affects the estimate, which is now about 0.04 standard deviations (column 3).[25] The estimates in columns 2 and 3 are only weakly significant.[26] I also estimate equation 1 for those aged 60 years or more in 1981. These estimates are shown in row 1 of columns 4-6. The estimate in column 4 is significant, but when more variables are added it decreases considerably and is no longer significant. The estimates for this old age group appear to be more sensitive to the inclusion of control variables and should therefore be interpreted with some care.[27]

Next I estimate OLS and IV regressions of health in 1981 on average income in 1967-1981 and other covariates. I have tried several functional forms for income, and the data strongly preferred income expressed in logarithmic units. Thus, I also follow much of the previous literature on health and income. The equations to be estimated are:

$$(2) \qquad H_{i81} = a + b \log(\bar{I}_{i81,15}) + q'X_{it} + e_{i81}$$

$$(3) \qquad \log(\bar{I}_{i81,15}) = p_0 + p_1 \bar{L}_{i81,13} + t'X_{it} + v_{i81},$$

where H_{i81} is the standardized index of bad health (STDH) for individual i in 1981; $\bar{I}_{i81,15}$ average income 1967-1981 (see section 2 for an exact definition); $\bar{L}_{i81,13}$ and X_{it} are defined above and; e_{i81} and v_{i81} are random error terms. Note that the magnitude of the estimated income effect, â, in equation 2 is easily interpreted. If b =-0.1, a 10 percent increase in income gives approximately 10-percent of a standard deviation increase in good health, on average.

The first three columns of row 2 of Table 6 show estimates, based on the estimation of equation 2 for all players, with less and more covariates. With only women and the cubic in age as controls, a 10-percent increase in income is estimated to generate around 0.05 standard deviations better health (column 1). Adding family background, socioeconomic variables and health at the beginning of the period has a large effect, reducing this estimated effect to 0.02 (column 2). I also estimate equation 2 for those aged 60 years or more in 1981. It appears that income offers more protection against poor health for older individuals. However, the estimates are not significantly different from the estimates using all ages.[28]

In addition to estimating health as a direct function of lottery prize, using OLS, I also conduct IV estimations where I use the average lottery prize 1969-1981 to predict log average income 1967-1981 (equation 3) which is then used in the second stage health regression (equation 2). The reason for these additional estimations is that I want to estimate health as a function of log income, since the data prefers the specification in log income and the usage of log income facilitates interpretation. The specification of income in logs is also in accordance with much of the earlier literature on income and health.[29] The IV estimates are shown in row 3 of Table 6. For all players, a 10-percent increase in income, only controlling for women and a cubic in age, is estimated to generate 8 percent of a standard deviation better health and this estimate is significant. Adding control variables decreases the effect to about 4-5 percent of a standard deviation. Note that the OLS effect using only women and a cubic in age (column 1 of row 2) is very close to the IV estimate with all exogenous control variables included (column 2 and 3 of row 4). The IV estimations for the oldest are very sensitive to the inclusion of control variables and the estimate is insignificant in columns 5 and 6. Note that the IV estimations scales the reduced form estimates to give these the interpretation of percentage increases and comparable to OLS:s preferred functional form for income in logarithmic units.[30]

The estimates of equation 3 (the first stage) are shown in row 4. There is a strong significant effect of lottery prizes on the logarithm of income, as is expected since the former is included in the calculation of the latter. A lottery winning of 10,000 SEK per year is estimated to generate about 8.5 percent higher income. This figure is in line with what is expected, given that the average income 1967-1981 for all players is 121,300 SEK (see Table 1), since 10,000 is 8.2 percent of 121,300. The first stage effect for those aged 60 or more is about 7 percent and also highly significant.

In Table 7, I estimate regressions using separate health outcomes as health measures. Note that in all columns in Table 7, I control for the same variables as before, except that instead of the standardized index of bad health in 1968, I control for the lagged dependent variable in 1968. In columns 1-6 this means controlling for a linear lagged health symptom variable, whereas in columns 7-10, I control for two dummies indicating some and severe headache/overweight. Since the number of symptoms is a count variable, I also estimate Poisson regressions in columns 2, 4 and 6. I also use Ordered probit models for headache and overweight since these are ordered variables (measured as no, some or severe). Note that the parameter estimates are not comparable using OLS and Poisson or Ordered probit models. In the first row of Table 7, I report estimates regressing the health outcome on the average lottery prize 1969-1981, controlling for other covariates. Higher lottery winnings are estimated to significantly improve mental health and reduce overweight. In row 2, health outcomes are regressed on average income 1967-1981 and covariates using OLS. I find statistically significant evidence of income being negatively associated with poor mental state, cardiovascular symptoms and headaches. Using IV (row 3), I find statistically significant evidence of income causally improving mental health and reducing overweight, just as in the reduced form estimations (row 1).

In estimating the effect of income on mortality, I estimate equations 1-3, but the dependent variable in equations 1 and 2 now being whether the individual dies within 5 or 10 years after the survey in 1981. The other variables are the same as when using STDH as dependent variable. Row 1 of Table 8 reports results from estimating equation 1, the reduced form, using a linear probability model. In row 1 of columns 1 and 2, an increase in yearly lottery winnings by 10,000 SEK is estimated to reduce the probability of dying within 5 or 10 years by 2-3 percent and these estimates are significant. For older individuals, this effect is estimated to 2-4 percent, that for dying within 10 years being significant. Row 2 of Table 8 reports results from estimating equation 2 by OLS. In row 2 of columns 1 and 2, a 10 percent increase in income is estimated to reduce the probability of dying within 5 or 10 years by 0.2-0.3 percent. For older individuals this effect is estimated to 0.2-0.8 percent. However, none of these estimates are statistically significant. In row 2, I instead use IV. The estimated effects are now about 10 times higher than those using OLS: A 10 percent increase in income is estimated to reduce the probability of dying by 2-3 percent and for older individuals, this estimate is 3-6 percent. Using lottery winnings as exogenous variation in income gives results suggesting income to be very protective against death: I believe the reduced form and IV point estimates to be too large to be credible. Note, however, that more reasonable effects are within the 95 percent confidence interval.[31]

4.2. Sensitivity analysis and comparison to other studies

Since few individuals have won very large prizes, my results could be driven by a small number of influential observations. I therefore repeated the regressions underlying the estimates reported in tables 6 and 8, by imposing lower prizes on those 38 individuals with total lottery prizes of SEK 100,000 or more (which is equivalent to yearly lottery prizes from 7,997 SEK to 100,728 SEK). This was done in two ways. First, by replacing these individuals' lottery prizes by the mean of the lottery prizes for this group (20,237 SEK) and second by changing their lottery prizes to the lowest of these 38 individuals' winnings (7,997 SEK). The results following this data transformation remain very similar. For instance, in the first (second) case, the estimate in column 3 of row 1 in table 6, changes from -.037 to -.042 (-.036), and the estimates in columns 1 and 2 of row 1 in Table 8 change from -.020 and -.029 to -.075 (-.027) and -.062 (-.022), respectively.[32]

The IV estimates in most of the estimations are larger (in absolute value) than the OLS estimates. This might seem somewhat odd (although this was also found in Ettner, 1996), since the OLS estimates, if biased, are likely to be upper bound estimates, i.e. the protective effect of income is overestimated.[33] The reasons are that variables that are difficult to perfectly control for, such as risk attitude, rate of time preference, genes and family background, are likely to bias the OLS estimate in this direction. The intuitive reason is that talents or traits leading to wise such investments (i.e. education and training) increasing permanent income, are likely to also lead to wise investments in health, for instance through wiser health behaviors (see Fuchs, 1982).[34] A common reason for an estimate being attenuated is measurement error. However, since I use income measures from registers which are averaged over many years, both measurement error and transitory variation in income should be of little concern. One reason for too low OLS estimates when control variables are added (see row 2 of Table 6) might be that these controls are measured with considerable error. Even if income is perfectly measured, but mismeasured control variables are included in the regression, the upper bound argument is no longer valid since measurement error in any of these control variables will bias the effect of income on health in an unknown direction. Furthermore, note that the OLS and IV estimates are almost never significantly different from each other. In fact, the OLS estimate with only women and a cubic in age as controls gives an estimated effect of income on health similar to the IV estimate with a full set of controls (see Table 6).

The OLS health-income and mortality-income estimates using only players are estimated to be about 40 percent larger as compared to using the full sample (players and non-players), using the specification in column 2 of Table 6 (see Appendix 2). This means inferences for the whole population should be drawn with care, when the results are based on a sample of lottery players. From Appendix 2 I also note that using temporary (row 1) or more permanent measures of income (rows 2 and 3) has very little effect on the magnitude of the estimated health and mortality effects.

How should the magnitude of the estimated effects in this study then be interpreted? Are 0.04-0.05 standard deviations a large or a small effect? Ettner (1996) used two US data sets, each including a measure of self-assessed health, measured on a five-point scale, and estimated an Ordered probit model of the health measures on log income and a number of covariates. For one data set, the estimation showed that a 10 percent increase in annual income increased general health by about 0.01 standard deviations. The second data set showed that a 10 percent increase in monthly income increased general health by 0.02

standard deviations. Both income measures were measured once, and as the sum of respondent and spouse income. When IV-techniques were used, the estimated effects increased to 0.04-0.09 standard deviations. Hence, the IV results are similar to the effects found in my analysis for Sweden.[35]

Snyder & Evans (2002) estimated the effect of income on mortality for a sample of US males aged 65 or more. As exogenous variation in income they used that social security benefits were based on the date of birth for this sample: those born before January 1, 1917 received higher benefits than those born after that date. Surprisingly, they find the former group to have a higher mortality rate, even though they limit the groups to those born in the fourth quarter of 1916 and the first quarter of 1917. They explain this counter-intuitive result by showing that the low-benefit group increased post-retirement work and arguing that this can have positive health effects. Their results do not really contradict the health effects found for the old age group in this paper (the lottery estimates for this group are small and insignificant), but it does contradict the mortality effects.

Is it reasonable to find income to be protective against bad health and mortality in Sweden, with its extensive safety net and narrow income distribution? One reason might be that not only absolute but also relative income is of importance for health. A given amount of additional income improves the individual's relative position in the income distribution much more in Sweden than in US. Due to Sweden's narrow income distribution, a doubling of disposable income moves the individual from the median to the 99th percentile in our data. Hence, if the individuals' ranking in the income distribution were of importance, I would expect to find a positive effect of income on health in Sweden.

5. Conclusions

This paper has presented a new approach for analyzing whether the observed strong positive relations between income and measures of good health and life expectancy can be interpreted as causal effects. In order to estimate the causal effect of income on health, I exploited the fact that more than one-fifth of all individuals had won at least SEK 1000 on lotteries 1969-1981. Establishing evidence that the players have somewhat different characteristics I limited the sample to players. I find that winning 100,000 SEK on lotteries in a 13-year period (almost 8,000 SEK per year) increases general health by 3 percent of a standard deviation at the end of this period and decrease the probability of dying within 5 (10) years after the end of this period by 2 (3) percent.

I also estimated health and mortality as a function of income using OLS and IV techniques, the latter using only the part of the income variation due to variation in lottery prizes. Using OLS, the findings are that an increase in income by 10 percent increases general health by 2 percent of a standard deviation and decreases the probability of dying within 5 (10) years of the last survey date by 0.2 (0.3) percent. Using IV, the findings are that income causally generates 4-5 percent of a standard deviation better health and a 2-3 percent lower probability of dying. I note that the estimates using OLS and IV do not differ significantly, which is true for both general health and mortality. Looking at specific health symptoms and diseases I found that income causally generates fewer symptoms of poor mental health and decreases problems with overweight.

I also estimated separate effects for individuals aged 60 or more. The health income-estimate is insignificantly different (but smaller) as compared to the estimate for all ages. Hence, income is not more protective against bad health for older people.

References

Adler, Nancy E., Thomas Boyce, Margaret A. Chesney, Sheldon Cohen, Susan Folkman, Robert L. Kahn and S. Leonard Syme. 1994. "Socioeconomic Status and Health: The Challenge of the Gradient." *American Psychologist* 49(1): 15-24.

Björklund, Anders and Mårten Palme. 2002. "Income redistribution over the life cycle and between individuals - empirical evidence based on Swedish panel data, in *The New Economics of Rising Inequalities*, eds. D. Cohen, G. Saint-Paul & T. Piketty. Oxford University Press, CEPR.

Case, Anne. 2001. "Does Money protect Health Status? Evidence from South African Pensions." NBER Working Paper 8495.

Currie, Janet and B.C. Madrian. 1999. "Health, Health Insurance and the Labor Market." in *Handbook of Labor Economics*, eds. O. Ashenfelter, and D. Card.

Davidson, R. and J.G. MacKinnon. 1993. *Estimation and Inference in Econometrics*. New York: Oxford University Press.

Deaton, Angus S. and Christina H. Paxson. 1998. "Aging and Inequality in Income and Health." *American Economic Review* 88(2): 248-253.

Deaton, Angus S. and Christina H. Paxson. 2001. "Mortality, Education, Income, and Inequality among American Cohorts." in *Themes in the Economics of Aging*, ed. D. Wise, University of Chicago Press: Chicago.

Duflo, Esther. 2000. "Child Health and Household Resources in South Africa: Evidence from the Old Age Pension Program." *American Economic Review: Papers and Proceedings* 90(2): 393-398.

Eriksson, Robert, and Rune Åberg. 1987. *Welfare in Transition. A Survey of Living Conditions in Sweden 1968-1981*, Clarendon Press, Oxford.

Ettner, Susan L. 1996. “New Evidence on the Relationship Between Income and Health.” *Journal of Health Economics* 15(1): 67-85.

Fuchs, Victor, R. 1982. “Time preference and Health: An Exploratory Study,” in *Economic Aspects of Health*, ed. Victor R. Fuchs. Chicago: University of Chicago Press, 1982.

Gerdtham, Ulf and Magnus Johannesson. 1999. “New estimates of the demand for health: results based on a categorical health measure and Swedish micro data.” *Social Science & Medicine* 49: 1325-1332.

Gerdtham, Ulf, Magnus Johannesson, L. Lundberg, and D. Isacsson. 1999. “The Demand for Health: Results from New Measures of Health Capital.” *European Journal of Political Economy* 15: 501-521.

Grossman, Michael. 1972. “On the Concept of Health Capital and the Demand for Health,” *Journal of Political economy* 80: 223-255.

Imbens, Guido W., Donald B. Rubin, and Bruce I. Sacerdote. 2001. “Estimating the Effect of Unearned Income on Labor Earnings, Savings, and Consumption: Evidence from a Survey of Lottery Players.” *American Economic Review* 91(4) : 778-794.

Lindeboom, Maarten, France Portrait, and Gerard J. Van den Berg. 2002. “An Econometric Analysis of Major Events in the Life of Older Individuals,” *Health Economics* 11: 505-520.

Lindh, Thomas and Henry Ohlsson. 1996. “Self-employment and Windfall Gains: Evidence from the Swedish Lottery,” *Economic Journal* 106: 1515-1526.

Lundberg, Olle. 1991. “Causal Explanations for Class Inequality in Health – An Empirical

Analysis." *Social Science & Medicine* 32 (4): 385-393.

Manderbacka, K. 1998. *Questions on Survey Questions on Health*, Phd Dissertation: Swedish Institute for Social Research, Stockholm.

Pritchett, Lant, and Lawrence H. Summers. 1996. "Wealthier Is Healthier." *Journal of Human Resources* 31(4): 841-868.

Smith, James P. 1999. "Healthy Bodies and Thick Wallets: The Dual Relation Between Health and Economic Status." *Journal of Economic Perspectives* 13(2): 145-166.

Snyder, Stephen and William N. Evans. 2002. "The impact of Income on Mortality: Evidence from the Social Security Notch," mimeo.

Sundberg, Gun. 1997. *Essays on Health Economics* Ph.D. Dissertation: Department of Economics, Uppsala University.

Van Doorslaer, Eddy, et al. 1997 "Income related inequalities in Health: Some International Comparisons." *Journal of Health Economics* 16(1): 93-112.

Appendix 1: Construction of the standardized index of bad health (STDH)

In order to construct the standardized index of bad health, I need to find a way of weighting the contribution of each of the health symptoms to overall health. For this purpose, I use data from an additional SLLS, conducted in 1991.[36] The respondents were then asked the same health questions as in earlier waves and for the first time they, were also asked a direct question on their general health status: "How do you rate your general state of health?". Answers were given in three categories: Good (0), Bad (2), and somewhere in between (1). This measure has been shown to capture general health status very well (see Manderbacka, 1998).

I estimate a model of general health as a function of all health symptoms in 1991, for those aged 34-76 in that year. Since the general health variable is coded in three ordered categories, I use an Ordered probit model, that is, a latent index of bad health (h*) depends on observed health variables (x) in the following way: $h^* = g'x + e$, where h=2 if $m_2 \leq h^*$, h=1 if $m_1 \leq h^* \leq m_2$ and h=0 if $h^* \leq m_1$; x is a vector of 48 health symptoms (included as 81 indicators); h is subjective general health (where h=2 if the respondent's health is poor, h=0 if good, and h=1 if in-between) and $e \sim N(0,1)$. The cutoff points m_j:s and the vector g are estimated.

The 48 health symptom questions were also asked in earlier SLLS-surveys, so that the 81 indicators are also available in these waves.[37] For most symptoms, respondents were asked to state whether he/she had had this symptom to no, some or a severe degree in the 12 months prior to the survey. In those cases where less than 1 percent was coded in any of the two symptom categories (no problem for first category), the some and severe group were combined.

The produced symptom-estimates are then used as weights, as I linearly predict respondents' general health in 1981, based on individuals' symptoms in 1981 and the estimates: $\hat{h}^* = \hat{g}'x$, where the estimated parameters for 1991 are shown in Table A1. Hence, $\hat{h}^*$ is constructed from information on 48 health symptoms <u>and</u> their respective weights. This predicted bad health index was then standardized to have a mean of zero and a variance of one.[38]

As shown in Table A1, the most common symptoms were (slight or severe) cold, migraine, lower-back pain and pain in the shoulder area, in the last 12 months. The symptoms contributing most to overall bad health were (in order of importance) cancer, severe general tiredness, severe bronchitis/asthma (breathing problem), not being able to run 100 meters without problems, mental illness and severe pain in the shoulders.

Using STDH as a measure of general health has several advantages as compared to using separate health indicators or other available overall health measures. First, it facilitates the interpretation of the parameter estimates, since the effect of a one-unit change in one of the exogenous variables can be interpreted in standard deviation units. Second, STDH is superior to separate measures, since it is more general, and to the number of health symptoms, since it is based on information on these health symptoms, but where each symptom is weighted according to its contribution in explaining general health status. Third, the estimations are greatly simplified. Several of the separate symptoms and the other health measures require non-linear estimation techniques, either due to being ordered in few categories or having extremely skewed distributions. Fourth, there is often insufficient variation in separate health variables (too few non-zero observations) to provide reliable estimates.

There are also some potential disadvantages of using STDH as a measure of bad health. First, problems might arise if the contribution of health symptoms to overall health (i.e. the weights) has changed over time. One way of testing for this is to separately correlate STDH with the number of sickness symptoms for 1968, 1974 and 1981. The correlations are 0.766, 0.776 and 0.830, respectively. Hence, I conclude that there are some indications of the weights having changed somewhat, but not a great deal, over time. Second, the health symptoms in 1981 might imperfectly capture general health in 1991, due to the importance of some new symptoms for general health having emerged between 1981 and 1991 (an obvious example is HIV/AIDS). This is probably not a great problem though, as indicated by the respondents' answers in 1991 to questions of whether they had sicknesses or health problems not included in the questionnaire. 11.1 percent of the respondents added one symptom, and only 1.4 percent added two symptoms, which suggests that the most important sicknesses and symptoms were originally included.

Table A1: Estimated health weights

Health symptom	Estim. (st.err.)	Health symptom (cont.)	Estim. (st.err.)
IMMOBILITY		STOMACH/ INTESTINAL PROBLEM (cont.)	
Cannot walk 100m quite fast without probl. (=.075)	**.47 (.10)**	Constipated (some=.041)	.01 (.12)
Cannot run 100m without problem (=..236)	**.69 (.07)**	(severe=.015)	.13 (.19)
Cannot walk in stairs without problem (=.104)	.39 (.09)		
		Gastric ulcer (some or severe=.023)	.33 (.16)
PAINS			
Pain in shoulders (some=.221)	.21 (.07)	Haemorrhoids (some=.051)	.08 (.11)
(**severe**=.129)	**.58 (.08)**	(severe=.015)	.30 (.20)
Pain in Back/sciatica (some=.228)	.06 (.07)		
(severe=.148)	.45 (.08)	OTHER	
Stiff/pain in the joints (some=.174)	.15 (.07)	Genital discomfort (some=.022)	-.21(.18)
(severe=.114)	.27 (.08)	(severe=.011)	.09 (.22)
		Menstrual discomfort (some=.040)	-.03(.16)
CARDIOVASCULAR DISEASES		(severe=.030)	.09 (.16)
Pain in chest (some=.070)	.08 (.09)	Migraine, Headache (some=.373)	.02 (.06)
(severe=.030)	.26 (.14)	(severe=.107)	.09 (.09)
Heart weakness (some or severe=.034)	.25 (.13)	Cold, Influenza (some=.487)	-.11(.06)
		(severe=.165)	-.06(.08)
High blood-pressure (some=.104)	.24 (.08)	**Seeing/Eye problem not** (some=.049)	.33 (.11)
(severe=.022)	.26 (.15)	**improved by glasses** (severe=.031)	**.51 (.13)**
Varicose vein/ulcer (some=.068)	.25 (.09)	Hearing problem (some=.148)	.19 (.07)
(severe=.014)	.37 (.18)	(severe=.041)	.23 (.12)
Swollen legs (some=.077)	.01 (.09)	**Chronic bronchitis/asthma** (some=.037)	.21 (.12)
(severe=.023)	-.35(.16)	**(luftvägsinfektion)** (severe=.016)	**.70 (.20)**
Shortness of breath (dyspnoea) (some=.056)	.15 (.11)	Bladder/Prostate disorder (some=.036)	-.06(.12)
(severe=.023)	-.17(.19)	(severe=.019)	.25 (.17)
Dizziness (some=.093)	.07 (.08)	Heart attack, coronary	.24 (.21)
(severe=.021)	.10 (.16)	(some or severe=.011)	
POOR MENTAL HEALTH		Struma (some or severe=.027)	-.23(.14)
General tiredness (some=.179)	.35 (.07)	Kidney problem/stone (some=.012)	.11 (.22)
(**severe**=.050)	**.71 (.12)**	(severe=.012)	-.27(.22)
Sleeping problem (some=.108)	.19 (.08)	Overstrained (some=.053)	-.08(.11)
(severe=.042)	.31 (.13)	(severe=.010)	-.13(.24)
Nervousness/anxiety (some=.088)	.23 (.09)	Hot flushes (sweatings) (some=.062)	-.25(.11)
(**severe**=.029)	**.46 (.17)**	(severe=.022)	-.31(.17)
Depression (some=.039)	.25 (.12)	Cough (some=.182)	.02 (.07)
(severe=.023)	.24 (.18)	(severe=.043)	-.12(.13)
Mental illness (some or severe=.010)	**.60 (.23)**	Anemia (blodbrist) (some or severe=.022)	.12 (.17)
STOMACH/ INTESTINAL PROBLEM		Inguinal hernia (some or severe=.017)	.17 (.18)
Stomach pains (some=.138)	-.03(.08)	Growing thin (unnatural weight loss)	.19 (.17)
(severe=.048)	.27 (.12)	(some or severe=.019)	
Gallstone(biliary colic)/bilious(some/severe=.030)	.07(.013)	**Cancer (some or severe**=.011)	**.86 (.20)**
Feel sick at one's stomach (some=.080)	-.10(.11)	Rashes/eczema/psoriasis (some=.101)	.06 (.08)
(severe=.019)	-.09(.20)	(severe=.021)	.12 (.16)
Vomiting/throwing up (some=.041)	-.13(.15)	**Diabetes (some or severe**=.032)	**.48 (.12)**
(severe=.016)	.28 (.22)		
Diarrhea (some=.076)	-.16 (.11)	Overweight (some=.128)	-.01(.08)
(severe=.019)	-.36 (.18)	(severe=.030)	-.12(.14)
		Organic nervous disorder (CP,MS, Polio etc) (some or severe=.009)	.32 (.23)

Notes: Number of observations is 3551, and the likelihood function is maximized at –1673. Estimations also included a cubic in age and a gender dummy. All individuals are between 34-76 years of age. The 10 indicators that best predict bad health are highlighted.

Endnotes

[1] See, for example, Smith (1999) for the US, and Lundberg (1991) for Sweden.

[2] Some recent studies analyzing the effect of income on health or mortality are Ettner (1996), Deaton and Paxson (1999), Case (2001), and Lindeboom et al. (2001). For recent evidence based on Swedish data, see Sundberg (1998), Gerdtham and Johannesson (1999) and Gerdtham et al. (1999). For an international comparison, see van Doorslaer (1997), and for a cross-country analysis, see Pritchett and Summers (1996). Also note that an increase in income can have an additional effect on the health of other family members (see Duflo, 2000, and Case, 2001).

[3] For a survey of the effect of health on wages, earnings and working hours, see Currie and Madrian (1999).

[4] Low income have a negative effect on health through several different mechanisms, for instance bad health behaviors, such as smoking and excessive consumption of alcohol, or reduced access to quality health care. Low income might also generate stress or depression or hostility. In addition, low income can lower income relative to others which can affect health. For a discussion of these explanations, see Adler et al. (1994) and Smith (1999).

[5] See Eriksson and Åberg (1987).

[6] Note that individuals above 76 years of age are not interviewed and hence, not kept in the sample. However, we still know if these individuals are alive, and if not, their date of death.

[7] The specific symptoms are listed in Table A1 in Appendix 1.

[8] Lindh & Ohlsson (1996) used the lottery information in 1981 as a dummy variable for whether the respondent ever won on lotteries, for analyzing self-employment and wealth.

[9] Note that at the time when the data used in this paper was collected, a prize was always paid out as a lump sum; either at the time of winning or within a couple of weeks.

[10] These results are available from the author upon request.

[11] For some individuals, this difference turns out to be negative. Since it is impossible for the sum of all previous lottery prizes to decrease over time, I put the sum of lottery prizes for these individuals to zero. Hence, these individuals are not included in the sample of players.

[12] Note that for the midpoints of these years, i.e. in 1971 (1978), SEK 5.79 (6.97)=$1 and in 1998, SEK 9.85=$1 according to OECD National Accounts PPP figures.

[13] The number of household members in the respondents' household equals 1+d+c, where d equals one if the respondent is married or cohabitant and zero otherwise; and c is the number of children aged below 18.

[14] For some sensitivity analyses I use two other income measures: Average income 1975-1981, calculated over the 7 most recent years (1975-1981) as $\bar{I}_{81,7} = \sum_{j=75}^{81} (y_j + \bar{L}_{81,7}) / a_j \Big/ 7$ and income in 1981, the most recent measure, calculated as $I_{81} = y_{81} / a_{81}$.

[15] The question concerned SEK 2,000 in 1968 (=SEK 13,464 in 1998) and SEK 2,500 in 1974 (=SEK 11,009 in 1998).

[16] Note that the maximum prize was SEK 500,000 in the lotteries in 1975.

[17] The number of individuals in Sweden in 1975, aged between 28 and 70 (since in 1981 they are between 34 and 76 in my sample), was 4,299,000, which means that my sample of 2,948 individuals is a sample of 0.07 percent of the population.

[18] Calculated as $(2{,}500 \cdot 626) + (0 \cdot 2{,}322)$

[19] In 1975, 260 MSEK out of the 550 MSEK were paid back to players on lotteries. The gambling figures are 280 out of 800 MSEK for betting on soccer. If we assume that 35 percent of the amount on horse betting went back to winners, this amounts to about 50 MSEK

out of 140 MSEK. Hence, a total of about 590 MSEK went back to the players on gambling and lotteries.

[20] The amount calculated based on SLLS is expected to be lower, since respondents were asked to report lottery winnings only if the total amount were at least 1000 SEK.

[21] The only exception is that non-players are more likely to suffer from headaches. But this is driven by gender differences, since if the test of mean equality in headache is done separately by gender, this is no longer significant.

[22] Note that the R^2 in Table 5 is based on the sum of squares, assuming constant error variances, whereas the p-values are from F-tests, where the error variances are allowed to differ (since robust standard errors are calculated for the regression estimates). Hence, the R^2-values and the p-values from the conducted F-tests are not exactly comparable.

[23] Controlling for lagged health is an approach supported by the "health capital"-theory (see Grossman, 1972), where health is a stock measure, and the current health stock equals the sum of the previous health stock (scaled with a depreciation term) and investments in health during the period. Hence, all control variables in estimation of a dynamic health equation will generate the current health stock by determining investment in health.

[24] Surprisingly, the standard errors on the lottery-effect estimate that are calculated without being robust to heteroscedasticity, are larger than the robust standard errors. I believe the robust standard errors reported in the tables to be more correct since homoscedasticity can be rejected with a Breusch-Pagan test and since using an alternative heteroscedasticity-robust version less sensitive to small samples (known as HC3, see Davidson and MacKinnon, 1993), these standard errors are very similar to the robust standard errors reported in the tables.

[25] Note that the estimates for other covariates in row 1 and column 3 of Table 3 give that worse health is significantly associated with being older, having a low education and having bad health in 1968.

[26] Regressing health directly on lottery prizes (the reduced form), using the specification in column 2 of Table 6, the estimate (standard error) is -0.045 (0.029). If I instead estimate health on predicted total income, where total income is predicted from lottery prizes in the first stage (i.e. using income and not log income in the first stage), the estimate becomes somewhat smaller compared to the reduced from (which is also the numerator in the IV estimate). It changes to -0.034 (0.029). The reason for the difference is that lottery prizes have a positive (although insignificant) effect on non-lottery income.

[27] We also performed some sensitivity analysis by using number of health related symptoms as dependent variable instead of STDH, using the same specification in column 3 of table 6. The estimates (standard errors) for lottery prizes were then -.071 (.138) using OLS and -.020 (.032) using a Poisson model. Hence, the t-values are lower compared to using STDH as dependent variables. The drawback with using the number of symptoms as dependent variable is that, for instance, cough gets the same weight as cancer. I therefore as dependent variable also used the number of serious health conditions, where a serious health condition is defined as being among the 10 best predictors of bad health in table A1 in Appendix (these are highlighted in the table). The lottery prize estimate (standard error) are then -.048 (.034) using OLS and -.131 (.095). These t-values are similar as the one when STDH is used as dependent variables. I therefore conclude that the weighting of health symptoms matters for the results whereas using the STDH weights or other weights do not matter for the results.

[28] There is also never a significant difference in the income estimates for women and men.

[29] Replacing the log of total income (including lottery prizes) by the lottery prize in equation 2, would not identify the percentage effect of total income on health, since I would estimate $b*p_1$ instead of b. b would also not be identified, since $\log(X+Y) \neq \log(X) + \log(Y)$, by regressing health on the log lottery prize.

[30] Running the same specification as in column 2 of Table 6, but replacing STDH with other outcome variables as dependent variables, gives that lottery winnings 1969-1981 are not significantly related to income in 1981 (estimate positive), average income 1967-1981 (estimate positive and smaller than above), the probability of working in 1981 (estimate negative), the hours worked in 1981 (estimate negative) and number of persons in the household in 1981 (estimate close to zero) and are significantly related to the probability of getting married or being a cohabitant in 1981 (estimate negative). Also, Lottery prize 1969-1981 has virtually no effect on the average equivalence scales 1967-1981. If the latter is regressed on the former, controlling for the same covariates as in column 3 of Table 6, the estimate (standard error) on lottery prize is 0.0001 (0.0198).

[31] If estimating the effect of lottery winnings on the probability of dying within 15 years the estimates change signs and become positive and insignificant for all players, and positive and significant for those aged 60 or more. I mainly focus on the 5- and 10-years results here, since it could be argued that 15 years is too long a time after the lottery winnings to expect any effect. Note that the individuals on average won prizes around 1975 and that 1996 is 15 years after the survey period. If I estimate the effect of lottery winnings on life expectancy for those who died within 15 years, the estimates are positive and very large. Lottery winnings amounting to 10,000 SEK yearly are estimated to increase life expectancy by six months and this effect is about the same for those aged 60 or more.

[32] If these 38 individuals with the highest lottery winnings are eliminated, the results in Table 6 are affected, the estimate in column 3, row 1 changes to .091, whereas the results in Table 8 are reinforced, the estimates in columns 1 and 2 of row 1 in Table 8 change to -.095 and -.086, respectively.

[33] That the IV estimations give consistent estimates of casual effects is strongly supported by the results in section 3.3 where most of the variation in lottery prizes was shown to be random. However, note that it is an underestimate of the protective effect of income if a risk loving, high time preference individual plays relatively more often (or for more money) on lotteries and thereby increases the chance of winning, and my control variables are not sufficient for conditioning on this. Note also that since winnings are likely to be from gambling as well as lotteries, it might be that intelligence is positively related to the amount of winnings. This means that if intelligence is not controlled for and intelligent people have better health, regressing health on lottery winnings could give an overestimate of the protective effect of income. Given that most winnings are from lotteries (see section 3.1.) and that it is probably doubtful how much intelligence is involved in winning at betting, this should not be a great problem.

[34] I here argue that risk-averse individuals are likely to have higher permanent incomes. Given that these individuals will make safer investments, this is probably fine. Note however that in Table 1, players have higher permanent incomes.

[35] The instruments in Ettner (1996) were state unemployment rate, work experience, parental education, and spouse characteristics. Note that the estimates are not exactly comparable to mine, also since income is measured at one point in time and not adjusted for family size (both which would likely bias the estimates downward). Furthermore, my calculations of the effect

sizes in that paper are based on the reported Ordered probit estimates, which probably makes them too large. Furthermore, her estimates are based on a representative sample whereas my estimates are only based on players.

[36] Unfortunately, the question about monetary lottery prizes was not included in SLLS 1991.

[37] Even though I have information on 50 health symptoms, only 48 are used since only 12 persons had tuberculosis and 25 women had birth complications in 1991, which was too small a number to provide any reasonable estimate on the effect of tuberculosis and birth complications on general health.

[38] This method is also used to linearly predict general health in 1968 and 1974. Note that the creation of the standardized bad health indexes was done for the relevant age groups for these previous years, and that these indexes are only included as controls in some regressions.

Table 1: Descriptive statistics

	All (n=2948)		Players (n=626)		Non-players (n=2322)		p-value
	Mean	St.dev	Mean	Stdev	Mean	Stdev	
Main health variables							
Standardized bad Health index (STDH) in 1981	0.00	1.00	.003	.997	-.001	1.001	.927
Nr of bad health symptoms (SYM) in 1981	5.65	4.77	5.51	4.71	5.68	4.79	.435
Nr of Immobility symptoms (IMSY) in 1981	.52	.96	.51	.96	.52	.96	.829
Nr of poor mental symptoms (MESY) in 1981	.55	.98	.54	.95	.55	.99	.664
Nr of Cardiovascular symptoms (CASY) in 1981	.75	1.18	.76	1.18	.74	1.18	.805
Overweight (OV) in 1981	.17	.42	.19	.45	.16	.41	.170
Headache (HA) in 1981	.49	.66	.41	.63	.51	.67	.000
Died within 5 years	.064	.244	.065	.248	.063	.244	.844
Died within 10 years	.139	.346	.131	.338	.141	.348	.521
Main Lottery variables							
Player=1	.21	.41	1	0	0	0	
Player before 1969=1	.09	.29	.18	.38			
Av. lottery prize 69-81, in 10t, ($\bar{L}_{81,13}$)			.25	.64	0	0	
Av. lottery prize 75-81, in 10t, ($\bar{L}_{81,7}$)			.31	1.10	0	0	
Av. lottery prize 69-74, in 10t, ($\bar{L}_{81,6}$)			.19	.56	0	0	
Main income variables							
Log average income 67-81: Log($\bar{I}_{81,15}$)	11.60 [11.53]	.34 [4.25]	11.65 [12.13]	.33 [4.52]	11.58 [11.37]	.35 [4.16]	.000
Log average income 75-81	11.64 [12.01]	.35 [4.52]	11.69 [12.63]	.34 [4.96]	11.62 [11.85]	.35 [4.38]	.000
Log income 1981	11.63 [12.08]	.38 [6.29]	11.66 [12.50]	.39 [6.40]	11.62 [11.97]	.38 [6.25]	.023
Demographic variables							
Age in 1981	53.31	12.26	53.23	11.67	53.33	12.42	.844
Gender: women=1	.51	.50	.35	.48	.56	.50	.000
Family Background variables							
Foreign=1	.05	.22	.04	.21	.05	.23	.361
Economic problems growing up=1	.27	.44	.26	.44	.27	.44	.513
Health problems when growing up=1	.21	.41	.23	.42	.20	.40	.158
Socioeconomic variables; 1968, 1974							
Years of schooling 1968	8.52	2.86	8.44	2.72	8.55	2.90	.389
Worker in 1968=1	.74	.44	.83	.38	.72	.45	.000
Married in 1968=1	.79	.41	.76	.43	.80	.40	.029
Low wealth in 1968=1	.29	.46	.32	.46	.29	.45	.170
Very low wealth in 1968=1	.14	.35	.13	.33	.15	.35	.174
Years of schooling 1974	8.87	3.14	8.92	3.14	8.86	3.14	.690
Worker in 1974=1	.74	.44	.83	.38	.72	.45	.000
Married in 1974=1	.81	.39	.78	.41	.82	.38	.037
Low wealth in 1974=1	.21	.41	.20	.40	.21	.41	.777
Very low wealth in 1974=1	.10	.30	.07	.26	.10	.31	.009
Other covariates							
STDH in 1974	0.00	1.00	.01	1.01	-.00	1.00	.717
STDH in 1968	0.00	1.00	-.02	.95	.01	1.01	.534
SYM in 1974	5.32	4.47	5.20	4.37	5.35	4.49	.457
SYM in 1968	4.57	4.15	4.34	4.13	4.64	4.15	.114
Log income in 1974	11.55	.56	11.61	.49	11.53	.57	.001
Log income in 1973	11.48	.49	11.54	.47	11.46	.50	.000
Log income in 1967	11.28	.59	11.30	.64	11.27	.58	.391

Notes: Log average income is the logarithm of yearly average disposable family income in adult equivalencies. For the income measures averaged over 1967-81 and 1975-81, lottery prizes are included. The absolute values of income measures are shown in brackets in SEK 10,000 units. All income measures are expressed as yearly averages in 1998 prices. In 1998, SEK 9.85=$1. For three individuals, who have missing observations for disposable family income in 1973, I replace these missing values by the average of their disposable family incomes in 1967 and 1974. Av. lottery prizes are the average lottery prize won and are expressed as yearly averages in SEK 10,000 in 1998 prices. The p-values are from a t-test of equality of means for the sample of players and non-players, allowing for unequal variances among the two groups.

Table 2: Correlation matrix for health and mortality measures. Full sample.

	STDH	SYM	DOC	IMSY	MESY	CASY	OV	HA	D5	D10
Standardized index of bad health in 1981 (STDH)	1									
Number of sickness symptoms in 1981 (SYM)	.83	1								
Number of visits to the doctor in 1981 (DOC)	.41	.37	1							
Number of Immobility Symptoms (IMSY)	.80	.60	.31	1						
Number of poor mental symptoms (MESY)	.65	.68	.31	.35	1					
Number of Cardiovascular symptoms (CASY)	.67	.74	.28	.50	.41	1				
Overweight (OV)	.19	.31	.07	.17	.12	.23	1			
Headache (HA)	.15	.34	.06	.01	.21	.12	.09	1		
Died 1981-1986 (D5)	.24	.18	.14	.26	.13	.18	.01	-.05	1	
Died 1981-1991 (D10)	.30	.21	.13	.31	.15	.25	.02	-.06	.65	1

All p-values<0.001 (from test of no correlation among two variables), except the correlation between overweight and D5 (p=0.52), Overweight and D10 (p=0.27) and headache and D5 (p=0.01). Number of observations is 2,948.

Table 3: Gambling and lottery playing in Sweden 1969-1981. Amounts in 1,000,000 SEK, current prices.

	Gambling (1)	Lotteries (2)	Total (3)
1969	360	360	720
1975	940	550	1,490
1981	1,860	2,220	4,080

Note: The gambling figures are from betting at soccer (stryktips) and horse racing (V65). The lottery figures are from LOTTO and Penninglotteriet. V65 were started in 1974 and LOTTO in 1980. Note that in the gambling figure, betting at the horse tracks is not included in the numbers. Including these would likely only have a minor impact on the betting numbers. Otherwise, these were the main gambling and lottery games available in Sweden during these times. All numbers are from annual financial reports.

Table 4: Linear probability regressions of player status (player=1) on individual characteristics. Full sample.

	(1)	(2)
Women=1	-.138	-.119
	(.015)	(.017)
Age	-.003	.019
	(.043)	(.044)
Age^2/k_1	.025	-.020
	(.087)	(.088)
Age^3/k_2	-.022	.000
	(.045)	(.045)
Foreign=1		-.022
		(.031)
Health problems when growing up=1		.028
		(.019)
Economic problems when growing up=1		-.028
		(.017)
Years of Schooling in 1968		-.003
		(.003)
Married in 1968=1		-.042
		(.020)
Worker in 1968=1		.023
		(.018)
Low wealth in 1968=1		.019
		(.017)
Very low wealth in 1968=1		-.005
		(.022)
Standardized index of bad health in 1968		.014
		(.008)
Log Income in 1967		-.000
		(.015)
Player before 1969=1		.205
		(.031)
P-value from partial F-tests: 1	---	.000
2		.012
R^2	.032	.038

Constant included in all estimations. Robust standard errors are shown in parentheses. Number of observations is 2948. Note that k_1=2*Mean(Age) and k_2=3*Mean(age^2) are constants taking on values such as the marginal effect of age on the expected value of the dependent variables, for the mean individual, can be calculated by simply summing over the age coefficients. OLS is used in both columns. The p-values are from partial F-tests, where the null hypothesis in 1 is that all parameters, except the one for women, are equal to zero. The null hypothesis in 2 is that all parameters except the ones for women and player before 1969, are equal to zero.

Table 5: OLS/Tobit regressions of average lottery prizes 1969-1981 (in SEK 10,000) on individual characteristics. The sample consists of players.

Dependent variable:	Average lottery prize 1969-1981	Average lottery prize 1969-1974		Average lottery prize 1975-1981	
	OLS	OLS	Tobit	OLS	Tobit
			dy/dx		dy/dx
	(1)	(2)	(3)	(4)	(5)
Women=1	-.10	-.12	-.17	-.02	-.11
	(.03)	(.04)	(.09)	(.07)	(.10)
Age	.16	.41	.72	.06	-.10
	(.15)	(.16)	(.32)	(.28)	(.35)
Age^2/k_1	-.36	-.83	-1.45	-.20	.12
	(.29)	(.32)	(.63)	(.57)	(.72)
Age^3/k_2	.20	.42	.74	.14	-.02
	(.15)	(.16)	(.32)	(.31)	(.38)
Foreign=1	.03	.08	.06	-.04	.05
	(.06)	(.11)	(.21)	(.08)	(.17)
Health problem growing up=1	-.10	-.07	-.13	-.13	-.17
	(.04)	(.04)	(.10)	(.07)	(.11)
Economic probl. growing up=1	-.01	.02	.05	-.05	-.08
	(.05)	(.04)	(.09)	(.09)	(.13)
Years of schooling in 1968	-.00	.01	.03		
	(.01)	(.01)	(.02)		
Married in 1968=1	.01	-.05	-.12		
	(.05)	(.06)	(.11)		
Worker in 1968=1	-.09	-.09	-.14		
	(.05)	(.05)	(.11)		
Low wealth in 1968=1	-.01	-.05	-.21		
	(.05)	(.05)	(.11)		
Very low wealth in 1968=1	-.09	-.06	-.27		
	(.03)	(.04)	(.13)		
Standardized index of bad health in 1968	-.02 (.02)	.00 (.02)	.01 (.05)		
Years of Schooling in 1974				-.02	-.04
				(.01)	(.02)
Married in 1974=1				-.06	-.04
				(.11)	(.14)
Worker in 1974=1				.17	.13
				(.15)	(.18)
Low wealth in 1974=1				.02	.08
				(.10)	(.14)
Very low wealth in 1974=1				.01	.03
				(.09)	(.15)
Standardized index of bad health in 1974				-.02 (.03)	-.04 (.04)
Log income in 1974				.10	.05
				(.08)	(.11)
Log income in 1973				.08	.13
				(.07)	(.11)
Log income in 1967	.01	-.05	-.16	-.01	-.02
	(.04)	(.03)	(.07)	(.05)	(.09)
Player before 1969=1	.19	.19	.34	.19	.17
	(.10)	(.09)	(.15)	(.19)	(.23)
Average lottery prize 1969-74				-.05	-.68
				(.08)	(.48)
P-value from partial F-test	.196	.310	.444	.865	.929
R^2	.032	.057	---	.023	---

Constant included in all estimations. Robust standard errors are shown in parentheses. Number of observations is 626. Note that k_1=2*Mean(Age) and k_2=3*Mean(age^2) are constants taking on values such as the marginal effect of age on the expected value of the dependent variables, for the mean individual, can be calculated by simply summing over the age coefficients. The p-value is from a partial F-test, where the null hypotheses is that all parameters, except the one for women, are equal to zero.

Table 6: Regressions of the standardized index of bad health in 1981 on average lottery prize 1969-1981 (in SEK 10,000) and log average income 1967-1981. The sample consists of players.

	All players (n=626)			Players: Age>=60 (n=204)		
	(1)	(2)	(3)	(4)	(5)	(6)
Main independent variable	Dependent variable: STDH in 1981					
Average lottery prize 1969-1981 (the reduced form)	-.069 (.029)	-.045 (.029)	-.037 (.029)	-.086 (.043)	-.036 (.040)	-.016 (.044)
Log average income 1967-1981 (OLS)	-.51 (.12)	-.22 (.12)	-.22 (.12)	-.80 (.22)	-.33 (.23)	-.34 (.23)
Predicted log average income 1967-1981 (IV)	-.78 (.42)	-.53 (.40)	-.43 (.39)	-1.18 (.59)	-.54 (.58)	-.22 (.61)
	Dependent variable: Log average income 1967-1981					
Average lottery prize 1969-1981 (the first stage)	.088 (.023)	.085 (.019)	.086 (.020)	.073 (.021)	.066 (.019)	.070 (.020)
Controls:						
Women, cubic in age	Yes	Yes	Yes	Yes	Yes	Yes
X-variables in 1968 and STDH in 1968	No	Yes	Yes	No	Yes	Yes
Player before 1969	No	No	Yes	No	No	Yes

Constant included in all estimations. Robust standard errors are shown in parentheses. X-variables are the family background and socioeconomic variables listed in Table 1. STDH is the standardized index of bad health. In the first stage, log average income is regressed on average lottery prizes. In the reduced form estimations, STDH in 1981 is regressed on average lottery prizes. In the OLS estimations STDH in 1981 is regressed on log average income. In the IV estimations, STDH in 1981 is regressed on log average income which is instrumented by average lottery prizes. The mean (st.dev.) of STDH for all players are 0.00 (1.00) and for players age>=60 it is 0.51 (1.19)

Table 7: Regressions of separate health symptoms in 1981 on average lottery prize 1969-1981 (in SEK 10,000) and log average income 1967-1981. The sample consists of players.

Dependent variables:	Symptoms of Immobility		Symptoms of poor mental state		Cardiovascular symptoms		Headache		Overweight	
	OLS/IV (1)	Poisson (2)	OLS/IV (3)	Poisson (4)	OLS/IV (5)	Poisson (6)	OLS/IV (7)	Oprob (8)	OLS/IV (9)	Oprob (10)
Main independent variable										
Average lottery prize 1969-1981 (the reduced form)	-.057 (.040)	-.163 (-.180)	-.061 (.027)	-.283 (.144)	.063 (.065)	.070 (.056)	-.033 (.024)	-.129 (.102)	-.034 (.019)	-.212 (.137)
Log average income 1967-1981 (OLS)	-.17 (.11)	-.29 (.21)	-.23 (.12)	-.46 (.20)	-.34 (.13)	-.42 (.18)	-.14 (.08)	-.29 (.18)	.04 (.06)	.18 (.20)
Predicted log average income 1967-1981 (IV)	-.65 (.48)	-1.87 (2.08)	-.70 (.31)	-3.26 (1.69)	.73 (.71)	.74 (.55)	-.39 (.27)	-1.50 (1.17)	-.39 (.23)	-2.52 (1.61)

Constant included in all estimations. Robust standard errors are shown in parentheses. In all columns controls for women, cubic in age, the health symptom in 1968 (included as a continuous variable in column s(1)-(6) and as dummies in columns (7)-(10)) and all X-variables (family background and socioeconomic variables listed in Table 5) and an indicator for player before 1969 are included. In columns 2, 4 and 6, I estimate Poisson models and in columns 8 and 10, I estimate Ordered probit models. In the IV estimations we estimate a Poisson/Oprobit of health symptom on log average income, controlling for the first stage residual (from a regression of log average income on average lottery prizes).

Table 8: Regressions of dying within 5 and 10 years after 1981 on average lottery prizes 1969-1981 (in SEK 10,000) and log average income in 1967-1981. The sample consists of players.

	All players (n=626)		Players: Age>=60 (n=204)	
Dependent variables:	Dead in 5 years	Dead in 10 years	Dead in 5 years	Dead in 10 years
Main independent variable				
Average lottery prize 1969-1981 (the reduced form)	-.020 (.007)	-.029 (.011)	-.023 (.020)	-.042 (.017)
Log average income 1967-1981 (OLS)	-.023 (.031)	-.032 (.039)	-.020 (.091)	-.080 (.111)
Predicted log average income 1967-1981 (IV)	-.227 (.078)	-.334 (.129)	-.324 (.282)	-.594 (.245)

Constant included in all estimations. Robust standard errors are shown in parentheses. In all columns controls for women, cubic in age, STDH in 1968 and all X-variables (family background and socioeconomic variables listed in Table 5) and an indicator for player before 1969 are included. Note that the first stage estimates are the same as those in row 4 of Table 6. The mean (st.dev.) of Dead within 5 years for all players are 0.07 (0.25) and for players age>=60 it is 0.15 (0.36). The mean (st.dev.) of Dead within 10 years for all players are 0.13 (0.34) and for players age>=60 it is 0.30 (0.46).

Appendix 2: OLS regressions of standardized bad health index in 1981 (STDH in 1981) and dying within 5 years of survey in 1981 on log income. Less and more permanent income measures. Full sample.

Dependent variable	Standardized index of bad health in 1981						Dead within 5 years			
Sample	All (n=2948)				Age>=60 (n=1031)		All (n=2948)		Age>=60 (n=1031)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Log income in 1981	-.35	-.17	-.11	-.10	-.50	-.11	-.021	-.011	-.058	-.030
	(.05)	(.05)	(.04)	(.04)	(.11)	(.09)	(.011)	(.011)	(.032)	(.034)
Log average income 1975-1981	-.42	-.21	-.13	-.09	-.65	-.22	-.024	-.013	-.056	-.026
	(.05)	(.05)	(.05)	(.05)	(.10)	(.09)	(.013)	(.014)	(.031)	(.035)
Log average income 1967-1981	-.46	-.24	-.14		-.66	-.22	-.027	-.015	-.058	-.028
	(.05)	(.06)	(.05)		(.10)	(.10)	(.014)	(.015)	(.030)	(.034)
Controls:										
Women, cubic in age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
X-variables in 1968	No	Yes	Yes	No	No	Yes	No	Yes	No	Yes
X-variables in 1974	No	No	No	Yes	No	No	No	No	No	No
STDH in 1968	No	No	Yes	No	No	Yes	No	Yes	No	Yes
STDH in 1974	No	No	No	Yes	No	No	No	No	No	No

Constant included in all estimations. Robust standard errors are shown in parentheses. X-variables are the family background and socioeconomic variables listed in Table 1. STDH is the standardized index of bad health. The mean (st.dev.) of STDH in 1981 for all individuals are 0.00 (1.00) and for those age>=60 it is 0.50 (1.17). The mean (st.dev.) of Dead within 5 years for all individuals are 0.06 (0.24) and for those age>=60 it is 0.14 (0.35).