DISCUSSION PAPER SERIES

IZA DP No. 442

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

Mikael Lindahl

February 2002

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

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

Mikael Lindahl
University of Amsterdam and IZA, Bonn

Discussion Paper No. 442
February 2002

IZA

P.O. Box 7240
D-53072 Bonn
Germany

Tel.: +49-228-3894-0
Fax: +49-228-3894-210
Email: iza@iza.org

IZA Discussion Paper No. 442
February 2002

ABSTRACT

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

A vast literature has established a strong positive association of income with health status and a negative association with mortality. This paper studies the effects of income on health and mortality, using only the part of income variation that is due to a truly exogenous factor: the monetary lottery prizes of individuals. The findings are that higher income causally generates good health and that this effect is of similar magnitude as when traditional estimation techniques are used. A 10 percent increase in income increases good health by about 0.01-0.02 standard deviations.

JEL Classification: D31, I12

Keywords: Health, mortality, income, lottery

Mikael Lindahl
Department of General Economics/NWO 'Scholar'
Faculty of Economics and Econometrics
University of Amsterdam
NL-1018 WB, Amsterdam
The Netherlands
Tel.: +31-20-525 43 12
Fax: +31-20-525 43 10
Email: mlindahl@fee.uva.nl

* This work was begun when I was affiliated with the Swedish Institute for Social Research, Stockholm University. I thank Anders Björklund, Anne Möller Dano, Edwin Leuven, Olle Lundberg, Maria Melkersson, Hessel Oosterbeek, Helena Persson, Erik Plug, Katarina Richardson, Marianne Sundström, Bas van der Klauuw, Eskil Wadensjö, and seminar participants at EALE in Regensburg 1999, ESPE 2000, and SOFI for valuable comments and discussions. I also thank HSFR for financial support.

1. Introduction

It is well known that individuals with a 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, a change in any policy affecting people's income must take into account the additional effect on health (since policy makers very likely care about both the income and the health of their citizens). If an increase in transfer payments or certain tax cuts, besides increasing income, for a certain group, also leads to better health, it would be an additional argument for this policy. Calculating the effect of policies, such as a cut in marginal tax rates which makes the income distribution more dispersed, the extended distribution of health outcomes must also be taken into account. Hence, cost-benefit analyses of all public policies affecting income must consider the additional effect of income on health.

This paper focuses on whether disposable income has a causal effect on health. There are three main reasons why we observe a positive association between disposable income, as well as other income measures, and good health. First, this might be due to a spurious association between income and health, driven, for instance, 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, that is, the effect of health on income, for example through bad health decreasing work productivity, working hours or increasing medical expenditures, all of which would reduce the disposable income.[2] Third, it might be due to the causal effect of income on health.[3] If such an effect exists, the next step is to investigate through which channels it operates. If there is no causal effect, these explanations can be disregarded, and we 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.

A low income could affect 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. It might also generate stress, or create psychological states, such as depression or hostility. In addition, relative rather than absolute income could be of importance for health. This study does not attempt to sort through these explanations, however;[4] instead, I examine whether this causality exists at all.

Simply relating health variables to income measures, while controlling for other variables, is likely to be insufficient for consistently estimating the causal effect of income on health. The reverse causality issue discussed above or insufficient existing variables for which to control create a need for alternative identification strategies. This study estimates the causal effect of income on health by 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

---

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

[2] For a survey of the effect of health on wages, earnings and working hours, see Currie and Madrian (1999). Further, note that bad health can have a positive effect on current disposable income. This is the case if the marginal utility of consumption declines in poor health, since an individual with bad health would decrease her consumption and thus, increase her wealth and disposable income (see Smith, 1999).

[3] Some recent studies which have analyzed 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). Furthermore, note that an increase in income can have an additional effect on the health of other family members (see Duflo, 2000, and Case, 2001).

[4] For a discussion of these explanations, see Adler et al. (1994) and Smith (1999).

have the same chance to win (which is true if they all buy the same number of equally priced lottery tickets and the lottery drawing is correctly administrated), monetary lottery prizes create exogenous variation in income among individuals.

In this paper, I use data from the Swedish Level of Living Surveys (SLLS). SLLS consist of several waves and, in each wave, a representative sample of Swedes is interviewed, including those from earlier waves. Disposable family income, adjusted for the number of household members, is used as income measure. Hence, I attempt to measure the money available for spending for the interviewed person in the household. I have enough information, mostly taken from registers, to create an income measure spanning over 15 years but I also use less permanent measures of income. As exogenous variation in income, I use information on the amounts won on lotteries, taken from repeated surveys (SLLS), to which the register information on income has been matched. The SLLS also contain a vast number of questions regarding health symptoms, and register information on the death date of individuals. As health measures, I use a standardized index of bad health, constructed from the questions on health symptoms, and several measures of morbidity.

The next section discusses the data and some conceptual issues. In section 3, the mortality and morbidity measures are estimated as a function of income using traditional techniques. Section 4 investigates the exogenous nature of player status and monetary lottery prizes. In section 5, health and mortality measures are estimated as a function of income 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 (See Eriksson and Åberg (1987)). SLLS follow individuals across waves, so that many of the individuals are included in all years. New individuals are also added in each wave to maintain a representative sample.[5] A large 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.[6] All three waves of SLLS also contain a question on the amount of money won on lotteries. Below, I show how the health, mortality, lottery and income variables are constructed and I also show some descriptive statistics for the variables used.

<u>Health/Mortality</u>

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.[7]

---

[5] Note that individuals above 76 years of age are not interviewed and hence, not kept in the sample. The only thing we know about these individuals is whether they are alive, and if not, their date of death.
[6] I also have information on death dates and prior incomes for those who died between 1968 and 1981. This information is not exploited in the later analysis, however.
[7] For a detailed description of how STDH is constructed, see Appendix 1

Descriptive statistics for the number of bad health symptoms and STDH are shown in Table 1. The mean individual has 5-6 symptoms, and the distribution is left- skewed. STDH is standardized, with mean zero and standard deviation one, but somewhat left-skewed. For the mean individual, a one standard deviation increase in bad health is equivalent to a move from the 50^{th} to the 78^{th} percentile, which is equivalent to an increase in the number of health symptoms from 4 to 8.

Using STDH as a measure of general (bad) health has several advantages, compared to using separate health indicators or other available overall health measures, such as the number of health symptoms, visits to the doctor or weeks in bed due to sickness. First, different health measures are often used in different studies, which makes it hard to understand the magnitude of an estimated parameter and hence, to compare results across studies. Instead, my standardized index facilitates the interpretation of the parameter estimates, since the effect of a one-unit change in one of the exogenous variables in this index can be interpreted in standard deviation units.[8]

Second, STDH is superior as a measure of overall health, compared both to separate measures of health symptoms and to a measure of the number of health symptoms. STDH is superior to separate measures, since it is more general. STDH is superior 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. The sum of these symptoms is basically an unweighted measure of STDH (see appendix 1). 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

[8] A similar problem exists in the educational literature, where the effect of some treatment on test scores is often analyzed. Since test scores are based on different tests across studies, the scores are often standardized in order to be interpretable.

variation in the individual health variables (too few non-zero observations) to provide reliable estimates in itself.

There are, however, 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 implicit way of testing for this is to correlate STDH with the number of sickness symptoms separately 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 capture general health in 1991 imperfect, 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.

Due to these potential limitations, and because STDH has not previously been used in the literature, the sensitivity of the results from using this measure will be checked by using other measures of bad health as the dependent variable in estimating the effect of income on health. These variables are the number of health symptoms, the number of visits to the doctor, the number of weeks in bed due to sickness, and a couple of indexes capturing several related symptoms. These results will be reported in section 3.1.

The SLLS data has also been matched with register data on death dates for individuals deceased before January 1 1997 (i.e. within roughly 15.6 years from the survey year, 1981). By taking the difference between the death date (or the last day in December 1996 for those still alive) and the last interview date in 1981, I create a continuous variable capturing the number of Years left alive (YLA), which I use as a dependent variable in the estimations.[9] Since the individuals' age in 1981 is controlled for in the estimations, the estimate of the effect of income on YLA is exactly interpretable as the effect of income on life expectancy, since the sum of YLA and the age in 1981 equals the age at death. In Table 1, we see that, on average, respondents had almost another 14 years to live. However, only 24 percent died before 1997, so YLA is right- censored for the majority of individuals.

In Table 2, I correlate some health (measured in 1981) and mortality measures. In addition to the measures already discussed, we also use a variable capturing the number of visits to the doctor in 1981.[10] As mentioned above, STDH is highly correlated with the number of health related symptoms in 1981. STDH is more strongly correlated with the mortality measure and with the number of visits to the doctor, than what is the case for the number of symptoms measure. This indicates that the weighting scheme, in addition to reflecting the many different health symptoms, also captures other features associated with general health.[11]

To summarize, I use the following health variables: STDH in 1981 as a measure of the general health status and YLA as measure of mortality. 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. The measure of mortality is based on objective (i.e. based on register) information, however and, in addition, will capture some different aspects of health, i.e. those leading to death.

[9] Note that the register information on the date of death is not restricted to individuals aged 76 and younger. Hence, we know the death date of all individuals, even those who died after the age of 76.

[10] This variable has mean (standard deviation) equal to 2.28 (4.04).

[11] For the sub-sample of those who died before 1997, the correlation between STDH and the number of symptoms increases to 0.87, the correlation between the number of symptoms and visits to the doctor remains the same, whereas the correlation of YLA with both the number of symptoms and STDH decreases to -0.16 and -0.19, respectively.

Lottery Prize

In the SLLS for 1968, 1974, and 1981, the respondents were asked: "have you ever in your life won at least SEK 1,000 on betting or lottery of any kind?".[12] If this question was answered by '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 the respondent has won until the time of the survey. This information is then used to construct the lottery prize measures.[13]

Using the information from these lottery questions entails a couple of potential weaknesses. First, the second question above is somewhat unclear. How much the respondent has won altogether might mean the total sum at the time when he/she made the largest win. Interpreting the question in this way means that the lottery prize measure should have been constructed in a somewhat different way. However, using such an alternate lottery measure produces very similar results.[14]

Second, the SLLS did not contain a question of how often or how much the respondent plays on lotteries. This is potentially a great disadvantage, since people who play on lotteries likely play different amounts. If the respondent considered the question to be about 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 4 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

[12] Lindh & Ohlsson (1996) have previously used the lottery information in 1981 as a dummy variable of whether the respondent ever won on lotteries, for analyzing self-employment and wealth.
[13] Note that at the time when the data used in this paper was collected, a prize was always paid out on one occasion, which was either at the time of winning or within a couple of weeks.
[14] These results are available from the author upon request.

respondent has won since the previous survey.[15] Dividing the sum of lottery prizes by the number of years since the previous survey (i.e. by 6 and 7 respectively), I get the amount of yearly lottery prizes won between 1969-1974 and 1975-1981, expressed as yearly averages. In order to get a lottery prize in 1981 monetary value, I adjust the prizes using Statistics Sweden CPI figures for the midpoint between these years (July 1971, and December 1978) as base years.

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 on the lotteries since I know who stated 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, 26 percent of the full sample are players.

Are these potential weaknesses in the lottery information likely to seriously affect the results? Quite strong evidence that this is unlikely to be the case 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.

[15] For some individuals, this difference turns out negative. Because 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.

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 the lottery and non-lottery players will be made in section 4. There, I will also show more statistics for the lottery players, and conduct a detailed analysis of what determines player status and the amount of lottery prize won.

Income

The SLLS also have detailed information from tax registers (Statistics Sweden), which has been matched against the individuals. This includes the income from several different sources, such as income from work, capital and government transfers, and information on the amount of taxes paid (see Björklund & Palme, 2001, for details). This information is available from 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, mostly based on registers. Using this, and the lottery prize information for the periods 1969-1974 and 1975-1981, we 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 between year t-k and t, as $\overline{I}_{t,k} = \sum_{j=t-k+1}^{t} (y_j + \overline{L}_{t,k}) / a_j \Big/ k$. Family net income in year t, y_j, is calculated as tax-assessed income minus taxes plus transfers for the family (own and spouse), where the tax-assessed income includes pensions, sick pay, and unemployment benefits. Transfers include child and housing allowances; $\overline{L}_{t,k}$ 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, and the number of people in the household; a_j is the square root of the number of household members, where the number of household members is the number of children below 18, plus one if the respondent lives alone and plus two if the respondent is married or cohabitant; t=1974,1981 and k=6, 7.

In this paper, I use two measures of average income in the main analysis. . The first is calculated over 15 years, from 1967-81, 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. The second measure is calculated over the 7 most recent years (1975-1981) as $\bar{I}_{81,7} = \sum_{j=75}^{81} (y_j + \bar{L}_{81,7}) / a_j \Big/ 7$. In the later analysis, I also compare the health-income effect using these measures with the most recent (1981) measure, simply calculated as $I_{81} = y_{81} / a_{81}$. All these measures are expressed in 1998 SEK prizes.[16]

The reasons for using these different income measures are not only that I believe the possibly different associations of health with temporary and more permanent income to be interesting in their own right, but also because I want to compare the results for income, due its own variation, with income only due to the variation caused by lottery prizes. Since lottery prizes are temporary and the winners know this, it is important to be careful in comparing estimates. This relates to the discussion later in this paper on estimating lower and upper bound effects of income on health.

Descriptive statistics for the main income variables are shown in Table 1, with 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 if noting that doubling the average income for someone in the mid

[16] In 1998, SEK 9.85=$1 according to OECD National Accounts PPP figures.

50th percentile means that this individual moves to 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 being 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 that are constant over time for each individual. The average person in the sample is 53 years of age. The number of women is slightly lower than the number of men since men in Sweden on average die at a younger age than women. The share of foreigners, defined as people having immigrated at any age, is about 5 percent. . 27 percent grew up in families where the economic conditions were hard, 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 were quite similar, although the average number of years of schooling increased by about 3 months. In order to proxy for the wealth of the individuals, I use a question where respondents were asked to report whether they would have difficulties in bringing forth about SEK 12,000, in SEK 1998 value, within one week.[17] The responses are divided into three groups. Those individuals who were unable to bring

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

forth this amount are coded as being poor and thereby having "Very low wealth." Those individuals who could raise this amount of money either by loans or by some other way not including drafts from their own bank account, are probably individuals with a network of people with some wealth (such as family and friends) or some collateral to offer, but with no wealth of their own. I code them 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, income and lottery prize, which are used as controls in my later analysis to different degrees. I note that inequality in health, if measured by the standard deviation of number of sickness symptoms or the 90/10 percentiles for instance, seems to have increased over the years, whereas the inequality in disposable family income has decreased. The average individual has more sickness symptoms and a higher disposable family income. Since the sample contains the same individuals each year, these effects might just be due to them getting older, however.

3. Basic health and mortality regressions

Let us express the basic Health-Income relationship as:

(1) $$Health_i = \alpha + \beta f(Income_i) + \varepsilon_i$$

where *Health* is some measure of good health or life expectancy for individual *i*;[18] *Income* is some measure of the disposable income of individual *i*; ε is an error term that contains "everything else", i.e. both the characteristics the researcher can and cannot observe for the individual; α and β are parameters to be estimated.

[18] Later, I use a health measure expressed as an index of <u>bad</u> health. The example here is of good health, since this simplifies the discussion.

How to measure *Health* and *Income* was dealt with in section 2. The ideal functional form and whether a constant effect of *Health* on *Income* is a reasonable approximation will be touched upon in the next sub-section. Here, I instead ask the following important questions; what is hidden in the error term (and hence, for which variables would I like to be able to control), and which variables can be used as controls?

Variables that can safely be included as controls in equation (1) are those likely to have an effect on both *Health* and *Income*, but that are not themselves affected by *Health* or *Income* (at the time when these are measured). Candidates are variables capturing the respondents' demographics and family background and pre-determined socioeconomic, health and income and income variables. In the analysis below, I included a cubic in age, and indicator variables for women and for not being born in Sweden, as capturing demographics. I attempt to roughly capture early family background by two indicators for growing up in a family with health and economic problems, respectively. The socioeconomic variables are the number of years of education, work and marriage status, and two indicators capturing very low or low wealth of the individuals, all measured prior to *Health* and *Income*. Note that I do not include socioeconomic or health-behavioral variables measured simultaneously with (or after) *Income* and *Health*, since this would generate the risk of capturing income-health effects working through these variables as well as creating the problem that health itself potentially affects these variables (reverse causality).

Also including health measured at an earlier date as a control in the estimation- is an approach supported in "health capital"-theory (see Grossman, 1972, 2000). There, 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. Thereby, all variables (except the previous health stock) included as controls in a dynamic health equation will generate the current health stock by determining the investment in health.

I also include lagged income measures in some of the estimations. The reason for this is that when using lottery prizes as the sole variation in income, earlier income measures should not affect the estimate, if lottery prize is to be a good instrument for income. The lagged income measures also probably capture wealth effects to some degree. However, the estimate of the current averaged income variable should then be interpreted with some care, since the income variables are all highly correlated.

What then remains in the error term after including the variables just mentioned? Candidates are the degree of risk attitude and the rate of time preference (or the discount rate) and also factors such as genetics and family background. These characteristics might only be controlled for by my included covariates to some degree.[19] Further, note that controlling for lagged health measures does not necessarily solve these problems, since there could be many unobserved factors which do not only affect the health stock, but also the rate of investments in health.

For three reasons, I believe that the estimation of equation (1), including controls for the variables discussed above, is likely to estimate an upper bound of the casual effect of income on health. First, genetics and family background are very likely to favor the likelihood of earning high incomes and also the likelihood of better health. Hence, this mimics a classic omitted variables problem, leading to too high estimates of the effect of an explanatory variable, if the omitted variable is positively correlated with this variable and, net of this effect, with the dependent variable.

Second, I believe that an individual with a low time preference (and thereby a low discount rate), is likely to make relatively wiser investments in health (for instance through wiser health behaviors (see Fuchs, 1982)), but also those investments leading to a high

[19] I here disregard the sample selection problem, i.e. that those individuals with worse health are more likely to die earlier and hence, are more likely to have been lost from the sample.

permanent income (such as investment in training and education). If this is also true for risk averse individuals, both the unobserved risk degree and the time preference mimic the omitted variable situation discussed above. It should be noted, however, that costly investments which lead to a high permanent income create a lower current (temporary) disposable income. Hence, I could observe a different direction of this bias when using the current disposable income as an independent variable, compared to using more permanent measures of income.

Third, since I use income measure from registers (for most years) and these measures are also averaged over a number of years, I believe both measurement error and transitory variation in income to be of very little concern. Therefore, I believe that these factors do not bias my estimate (whereas if they existed, they would give an estimate biased toward zero).

Hence, I conclude that the (non-temporary) income effects estimated in this section are likely to be upper bound effects of the causal income effect. In section 5, I will argue that when my estimates only use the income variation due to variation in monetary lottery prizes, I am instead likely to estimate a lower bound.

3.1. Health regressions

In this section, I first estimate OLS-regressions of the standardized index of bad health in 1981 on average disposable family income (in adult equivalents) and other covariates, using a representative sample of Swedes.[20] I have tried several functional forms for income, and the

[20] Two previous studies of health in Sweden are Gerdtham and Johannesson (1999) and Gerdtham et al. (1999). The first used data from the 1991 wave of SLLS, and estimated an ordered probit model of categorical health on a number of covariates. The second study used another Swedish data set, and estimated three different measures of health on a number of covariates. Both these studies found that good health was positively affected by income. Despite these studies, there are several reasons for showing results using the full sample, and not just focusing on the lottery players. First, none of the previous studies have used the health or mortality measures used in this study, they did not use income measured over more than one year, and they did not control for previous health status. I therefore want to start by using a representative and as large as possible sample of Swedes. Second, I want to see whether the full (representative) sample and the lottery sample give similar health/mortality – income gradient estimates. Third, I want to test for functional form. Fourth, I want to test for whether the degree of permanence in the income measure used is of importance for the conclusions. Fifth, I want to test for whether there exist differences in the health/mortality-income gradient, by gender or age groups.

data strongly preferred income expressed in logarithmic units. Thus, I also follow much of the previous literature on health and income. The basic equation to be estimated is:

$$(2) \qquad H_{i81} = \alpha + \beta \log(\overline{I}_{i,k}) + \lambda X_{1,it} + \varphi X_{2,it} + \delta H_{it} + \varepsilon_{i81}$$

where $H_{i,81}$ is the standardized index of bad health (STDH) for individual i in 1981; $\overline{I}_{i,k}$ is the average income (see section 2 for an exact definition); $X_{1,it}$ is a vector of demographic (a cubic in age and an indicator for women) variables for individual i; $X_{2,it}$ consists of family background variables: indicators for being foreign, having had a bad economic situation and bad health in the family when growing up, respectively, and five socioeconomic variables for individual i: the respondents' number of years of schooling, indicators for work and marital status, and a proxy for wealth, measured as two dummies, of whether the respondent had low or very low economic status. The socioeconomic variables are measured as early as possible, i.e. in 1968 or 1974. In some of the estimations, I also include the previous health status (H_{it}), measured in 1968 or 1974. Whether the income coefficient is constant among groups, as assumed in equation (2), will be tested for below. Note that the magnitude of the estimated income effect, β, is easily interpreted. If β=-1, doubling the income gives a one standard deviation increase in good health, on average.

Rows 1-3 of Table 3 show estimates, based on the estimation of equation (2) for different income measures, with an increasing number of covariates. The income estimates from using income in 1981, averaged 1975-1981 or averaged 1967-1981, are surprisingly similar. Hence, whether temporary or more permanent measures of income are used only has a minor impact on the estimated effects. With only women and the cubic in age as controls, doubling the income is estimated to generate around a .4 standard deviation of better health. Adding family background and socioeconomic variables at the beginning of the period further reduces the income effect to about .2. Adding health at the beginning of the period has a large effect, reducing the estimate to just above .1. Altogether, the income estimate is very sensitive to the inclusion of exogenous variables in the estimation, which suggests that estimates of income-effects on health can be severely biased using traditional methods.

Rows 4-5 of Table 3 compare the income effect by gender. The income effect for women is insignificant, when health at beginning of the period is added. There is never a significant difference in the income effect for women and men. Note that the income effect for women is much more sensitive to the inclusion of previous health status as a control variable. In rows 6-9, I compare the income effects for four different age groups: 34-46, 47-60, 61-76 and 68-76. This effect is always stronger for the oldest individuals. In rows 11-12, the estimates of the interaction terms between income and age and income and gender are reported. I have expressed the estimates so that the main income estimates express the average income effects (in row 10), evaluated at the mean of age and women. Note first that these average income effects are very similar to the specification without interactions (row 3). The interaction terms for age are statistically significant and negative, which suggests that income offers more protection against poor health for older individuals. The interaction term between income and gender is always insignificant.

To check the sensitivity of the results, I also use other measures of bad health as dependent variables in the estimations. These variables are the number of sickness symptoms, the number of visits to the doctor, the number of weeks in bed due to sickness, and five indexes capturing several related symptoms (disability to move, tiredness, poor mental state, pains and cardiovascular diseases).[21] Applying the specification underlying the results in column 3 of Table 3, I find that all these health measures but one give a statistically significant effect of higher income being associated with better health. The only exception is the number of visits to the doctor, producing an insignificant estimate of higher income being associated with more visits to the doctor. However, we believe that this measure is likely to partly capture the fact that individuals with a higher income probably invest more wisely in health (by more often visit the doctor), which will produce better health. Even though visiting the doctor was inexpensive in Sweden in this period, the estimate might also partly reflect such costs. In appendix 2, it is shown that using the number of sickness symptoms as an outcome measure gives very similar results as using STDH. This is comforting, since it means that "the weighting"-procedure, used in constructing STDH, does not in itself produce adverse results. In appendix 2, I also show the estimates for the covariates underlying the income estimates in row 3 of Table 3. The results are that, controlling for permanent income, worse health is associated with being older, female, foreign, having had health or economic problems when growing up, having a lower education, being non-married, non-worker, and having had low or very low wealth.

3.2. Mortality regressions

In estimating the effect of income on mortality, I estimate an equation of the following form:

(3) $$YLA_i = \alpha + \beta \log(\bar{I}_{i,k}) + \lambda X_{1,it} + \varphi X_{2,it} + \delta H_{it} + \varepsilon_i \, ,$$

where YLA is the number of years left alive for individual i. Note that the effect of income on YLA, β, is exactly interpretable as the effect of income on life expectancy expressed in years (see section 2). The other variables are the same as before.

Table 4 reports results from estimating equation (4), using a Tobit. Using the different income measures gives similar results. A 10 percent increase in income is estimated to

[21] For a description of how these indexes were constructed, see Variabler and Koder foer LNU 1991.

increase life expectancy by 5-8 weeks.[22] This effect is very similar across gender groups, but differs between age groups. The estimates for the two lowest age groups are based on few non-censored observations, however. For the oldest individuals, for which more than half the observations are non-censored, I find that a 10 percent increase in income increases life expectancy by about 5 weeks. These effects are not statistically significant, however. Adding interaction terms to equation 4, an identical effect across both the gender and age groups cannot be rejected. As expected, when the fraction of uncensored observations is small, the Tobit estimates are much larger than the OLS estimates (which are not shown). Note, however, that the income effects are still positive for all groups, and significant for the average income effects, using OLS.[23] Regarding other covariates, life expectancy decreases with age and increases with being female, foreign, having been married and having good health in 1968 (see appendix 2).

In this section, I have presented upper bound estimates of the effect of disposable family income on health and life expectancy for a random sample of Swedes. An increase in permanent income by 10 percent increases health by 0.01-0.02 standard deviations and life expectancy by 5-8 weeks. I also found that permanent income is more protective against bad health for older people, whereas this was not the case for mortality.

[22] This is obtained by multiplying the estimates in columns 3-4 by 52/10.
[23] Note that regressing a dummy on whether the individual died before 1997 on permanent income gives the expected negative relationship, which is also significant.

4. Are playing on lotteries and the amount of monetary lottery prize won really exogenous?

4.1. Player-status

I start by contrasting the sample of players with that of non-players. This is done in Table 5, where descriptive statistics for both groups are shown. The last column contains p-values from a test of mean equality between the two groups. We see that players have, on average, significantly higher income and lower education. Players are also more likely to be older, a man, as well as having been single, worker and having had very low wealth previously.

Table 6 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 estimated to be much higher for men. Controlling for family background and pre-determined socioeconomic variables shows that the probability of playing increases with being foreign, having a lower education and having been single. The probability of playing also increases with bad health, but the estimate is not really significant (p-value=.14). Both lagged income and the proxies for wealth are insignificantly related to the probability of playing.

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 fourth of the original. However, there is also an intuitive reason for including only guaranteed 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 this 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 chose not to participate in the experiment. Notably, the results from the estimations using the lottery prize variable in section 5 remain basically unchanged if the whole sample is used.

4.2. Lottery prizes for the sample of guaranteed players

Returning to Table 5, we see that for the sample of players, the average yearly lottery prize is about SEK 2,000 (about $200) per year. This figure is probably comparable to a policy change in taxes and transfers of quite realistic magnitude. Note that this number corresponds to a total amount of lottery prizes of SEK 26,000 between 1968-1981. During this period, 151 individuals won no lottery prize, 305 individuals won positive lottery prizes of less then SEK 10,000 and 38 individuals won a lottery prize of more than SEK 100,000. All these figures are in 1998 year prizes.

In Table 7, monetary lottery prizes for players are regressed on the same covariates as in Table 6. Columns 1-2 use the monetary lottery prize during the whole period as a dependent variable. Using OLS, the lottery prize is significantly higher if the player is a man, had no health problem in the family when growing up, does not work (marginally significant), and had more than a very low economic status at the beginning of the period. The R^2 is not very high (0.017), even with this full set of variables.

Although all the individuals in these regressions are guaranteed to be players during 1969-1981, some (151 individuals) have not won any amount during this period. Those individuals who won no prize in this period, might have lost relatively more money. In that case, the lottery prize variable is censored just below the lowest positive value. In column 2, I therefore report estimates using a Tobit model. Then, only gender is significantly associated with lottery prize. It therefore seems that even women who play on the lottery play for smaller amounts. Note that, using both OLS and Tobit, we cannot reject that all variables, except women and the cubic in age, jointly have no effect on lottery prizes.[24]

Columns 3-9 of Table 7 present regression results separately by periods. Columns 3-4 show results for lottery prizes won in 1969-1974. Column 3 shows results from estimating lottery prizes on the full set of covariates using OLS. Lottery prizes are significantly higher if the player is a man, and significantly, but unsystematic, related to a cubic in age. The regression results also show that individuals previously earning a high income, won less lottery prizes. There is also some evidence that those with low wealth won less, and this effect is significant using Tobit.

Column 5 shows the estimates using the same covariates as in column 4, but now using the lottery prize in 1975-1981 as the outcome variable. Note that the average of the estimates in columns 3 and 5 equals the estimates in column 1. Hence, lagged income is now positively associated with lottery prizes in this second period. Columns 6-9 show results from estimating lottery prizes in 1975-1981 on the more recent set of covariates. Hence, I estimate an OLS regression of lottery prizes on demographic and family background characteristics, socioeconomic variables and health in 1974 and disposable family income in 1974. In column 7, I also add the income in 1967 and 1973. There is only weak evidence that lottery prizes are related to earning a high income in 1967-1974 (the p-value of a joint test for income 1967, 1973-1974 is 0.16).

In column 8, I add a variable for the amount of monetary lottery prize won in 1969-1974. The estimate reveals that the previous monetary lottery prize is unrelated to the current prize; a comforting result. Lagged lottery prize can be seen as a proxy for how much individuals play (or "the number of tickets bought"). Using a Tobit, this lagged lottery prize is

[24] Note that the R^2 in Table 7 is based on the sum of squares, assuming constant error variances, whereas the p-

negatively, but still insignificantly, related to lottery prizes won in 1975-1981. The insignificant effect of this variable suggests that people play for a similar amount all the time. If individuals plays for different amounts, we would expect this estimate to instead be positive, since those who always play for more money always would have a higher probability of winning.[25]

Altogether, the results in this section supports 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 only weak evidence that previous disposable family income is associated with the amount of lottery prize won, which suggests that "wealthier" people do not play for a higher amount and thus increase their chance of winning. This is also supported by the fact that the economic status indicators are not significant and that the number of years of schooling is significant but with a negative sign. 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 women and the age of the individual. Hence, the absence of information on the amount played seems to be a minor problem.

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.

[25] I also experimented with a fixed effect model, based on a dependent variable of the difference between lottery prizes in 1975-1981 and 1969-1974. All estimates did, however, turn out to be insignificant, and it is unclear whether this is due to the elimination of unobserved individual factors, the much smaller real variation that exists in variables measured as changes, or the small sample used here. Anyway, the estimated standard errors for the estimates in the FE-estimations were large enough not to produce significantly different effects as compared to the ones presented in Table 7.

5. Estimating causal effects of income on health and mortality

In section 3, using traditional techniques, I estimated an increase in permanent income by 10 percent to increase health by 0.01-0.02 standard deviations and life expectancy by 5-8 weeks. I argued, however, that these are likely to be upper bound effects, due to income being affected by a number of factors, unlikely to be captured by observable control variables. In section 4, I showed that most of the variation in monetary lottery prizes is due random variation, if the sample is restricted to those playing on lotteries. In this section, I therefore combine these findings, and estimate the effect of income on health using only the portion of income due to variation in monetary lottery prizes. Since I will only use the sample of lottery players from now on, I also show the income-health estimate using traditional techniques as a comparison.

In this section, I claim to estimate a causal effect or at least its lower bound. It is a causal effect if conditioning the sample to players and including the large number of control variables are sufficient to make lottery prizes a valid instrument for disposable family income. It is an underestimate if this is not enough and if a risk loving, high time preference individual is likely to play relatively more often (or for more money) on lotteries. This is due to the fact that risk-loving individuals, as shown in section 3.0, are likely to have both a lower permanent income and a lower current health stock. Still, in order to address these issues, I also add the variables capturing the amount of money previously won on lotteries. This should be positively correlated with the amount of money spent on playing on lotteries and thereby with the degree of risk likeness and the discount rate. Previous lottery winnings should be a strong control for these factors, at least if the degree of risk attitude and time preference are roughly constant for an individual during 1969-1981.

Formally, I estimate models of the form:

$$(5a) \qquad H_{it} = \alpha + \beta \log(\bar{I}_{it,k}) + \theta' X_{it} + \varepsilon_{it}$$

(5*b*) $\log(\bar{I}_{it,k}) = \pi_0 + \pi_1 \bar{L}_{it,k} + \tau' X_{it} + v_{it}$

where H_{it} is health for individual *i* at time *t* and $\bar{I}_{it,k}$ is the average (over k years) disposable family income for individual *i* at time *t*. Note that I use two measures of average income: one semi-permanent (75-81) and one more permanent (67-81) measure. X_{it} reflects the same control variables as used in section 3 and, in addition, lagged income and lottery prizes; $\bar{L}_{it,k}$ is the monetary lottery prize won at the same time as the income was earned, and ε_{it} and v_{it} are two error terms assumed to be uncorrelated.

Considering equation (5a), it might seem odd that I do not estimate health as a direct function of lottery prize, using OLS, instead of IV techniques. However, 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. Replacing the log of total income (including lottery prizes) with the lottery prize in equation (5a), would not identify the percentage effect of total income on health, since I would estimate $\beta * \pi_l$ (instead of β). β would also not be identified, since $\log(X+Y) \neq \log(X) + \log(Y)$, by estimating health as a direct function of the log lottery prize. Note, however, that if I estimate the standardized index of bad health as a direct function of the lottery prize, I obtain results that are basically identical (except for scale factor) as when using lottery prizes as an instrument for log total income, as specified in (5). The p-values are almost identical.

In Table 8, I show estimates of the effect of income on the standardized index of bad health, using players only. I show results for basic OLS estimates and IV estimates where the monetary lottery prize (between 1969-1981) is used as an instrument for income. Here, I also note that the (from the second stage) excluded lottery variable is always highly statistically significantly different from zero in the first stage (p-value<0.01).[26] The first row of columns 1-2 shows basic OLS estimates for this player sample. Compared to the full sample, the income estimates are now less sensitive to the inclusion of control variables and the effect is estimated to be 2-3 times larger. For instance, using the specification in column 2 of Table 8, I estimate the income effect to be -.33 using OLS and the sample of players, whereas this effect is estimated to be -.07 if only using the sample of non-players, and this result is only partly due to the larger fraction of men being players. This means that one should be careful in drawing inferences for the whole population, when the results are based on a sample of lottery players. The IV-estimates are very similar to the OLS-estimates, even though only the OLS estimates are significantly different from zero.

I also add an interaction term for income and age, where the instrument is lottery prize interacted with age. Note that I have scaled the main income effect so that it can be interpreted as an average income effect. The OLS estimates show that income is more protective against bad health for older people, whereas the IV interaction-estimates are close to, and not significantly different from, zero.

In columns 3-4, I repeat the estimations for the sample of players who won prizes between 1969-1981, for two reasons. First, it is possible that individuals who only won prizes before 1969might not have played on the lottery since then. Second, these individuals are, on average, older, since the probability of winning a prize increases with the number of times playing. Hence, the sample used in columns 3-4 is more representative of the whole sample of individuals. Similar results are found as in columns 1-2. The IV-estimates are now somewhat larger and also closer to being significant.

---

[26] This is not the case, however, if I use lottery prizes in 1969-1974 as instruments for average income in 1967-1974. The lottery variable estimate then has a p-value between 0.08 - 0.10 in the first stage regression.

In columns 5-7 of Table 8, estimates of the effect of income in 1975-1981 on the standardized index of bad health are shown. The lottery prize used as exogenous variation is now the prize between 1975-1981. Comparing the OLS and IV estimates, we see that, as before, the estimates are almost identical. However, the IV estimate now decreases a great deal, when control variables are added. When adding an interaction term for income and age, I see a different pattern emerges. Both the OLS and IV estimates reveal that income is more protective against bad health for older people, and the interaction effects are estimated with quite good precision.[27]

When using income, measured in 1975-1981, the data allows me to add several variables measured during an 8-year period prior to 1975. These variables are three income measures, and the amount won on lottery in 1969-1974. The health-income effect, as well as the interacted age-income effect, are unaffected by the inclusion of these variables. This supports the idea that no omitted factors (associated with risk attitude and time preference) bias the causal effect estimates.

Since few individuals have won very large prizes, my results might be due to a very small number of influential observations. I therefore repeated the regressions underlying the estimates reported in table 8, by putting identical weights on those individuals with yearly average lottery prizes of SEK 10,000 or more. This was done by changing these individuals' lottery prizes to the mean of the lottery prizes for this group. Hence, I down-weighted the very high lottery prizes, relative to the high lottery prizes. For the 34 individuals who won these high average lottery prizes in 1975-1981, their yearly prize ranged from SEK 11,252 to SEK 187,067. Since the mean yearly prize for this group is SEK 33,315, I replace their prizes by this number. This is done in a similar way for 1967-1981. The results following this data

[27] If regressions are only done for the sample of players with positive lottery prizes won in 1975-1981 (n=441), the results are very similar.

transformation remain very similar. For instance, the IV estimates in columns 2 and 4 of row 2 in table 8, change from -.39 to -.38, and from -.53 to -.56 , respectively.

Since the OLS and IV estimates in general are very similar in table 8, but about four times larger than those found using OLS for the full sample (table 3), I draw the following conclusions. First, an increase in income has a relatively higher beneficial health effect for individuals playing on lotteries, that is, income is of greater importance as predictor of health status for players as compared to non-players. Second, a 10 percent increase in income increases health by about .02 - .08 standard deviations for the sample of players, and there is some evidence that this effect is stronger for older people. Third, the causal effect of income on health is likely to be estimated fairly correctly by simple OLS techniques. Fourth, therefore, a 10 percent increase in income increases health by about .01 - .02 standard deviations for the full sample.

Are 0.01-0.02 standard deviations a large or a small effect? Ettner (1996) used two US data sets which each included 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 decreased general health by about 0.013 standard deviations. The second data set showed that a 10 percent increase in monthly income decreased general health by 0.021 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 3-5 times.[28] Nevertheless, the OLS results are remarkably similar to the effects found in my analysis for Sweden.[29]

[28] The instruments were state unemployment rate, work experience, parental education, and spouse characteristics.
[29] The estimates in Ettner (1996) are not exactly comparable to the ones in this paper, since the measures in that paper are measured at one point in time and not adjusted for family size (which would likely bias the estimates

Next, I show results from mortality regressions. Replacing H_{it} with YLA_i, the number of years left alive, in equation (5a), I estimate (5), using Tobit techniques. Columns 1-2 show results for income in 1967-1981, and columns 3-6 for income in 1975-1981. I first note that, once more, using the sample of players, the OLS estimates are about 2 times larger than the estimates using the whole sample. Using income in 1967-1981, I get unreasonably high IV estimates, pointing to a 10 percent increase in income increasing life expectancy by 6-8 months. For income in 1975-1981, I get much lower estimates. Furthermore, income does not appear to be more protective against mortality for older people. Since the dependent variable is heavily censored and the standard errors are very high in these estimations, I do not want to push for these results. Instead, I am looking forward to more research on the subject of income and mortality, where believably exogenous variation in income is used.

6. Conclusions and discussion

There is consensus among researchers that there is a strong positive association between measures of socioeconomic status and measures of good health and life expectancy. This paper presented a new approach for analyzing whether these correlations can be interpreted as causal effects. Estimating health and mortality measures as a function of disposable family income and some control variables, using traditional techniques and a random sample of Swedes, gives that an increase in income by 10 percent increases health by 0.01-0.02 standard deviations and life expectancy by 5-8 weeks. I do believe these to be upper bound effects of the effect of income on health and mortality. Further, I also found that income is more protective against bad health for older people, whereas there is an insignificant difference in the income-mortality estimate for people of different ages.

---

downward). Furthermore, the calculations of the size of the effects are based on the reported ordered probit estimates, which probably makes the calculated effect sizes too large.

In order to estimate the causal effect of income on health, I exploited that roughly one-fourth of the individuals had won at least SEK 1000 before 1981. Establishing evidence that the players have somewhat different characteristics, I limited the sample to players, and repeated the estimations for the whole sample. I then obtained the result that the income-health OLS estimates are 2-3 times higher than when using the whole sample of both players and non-players. Using monetary lottery prizes as an exogenous variation in disposable income for this restricted sample, I obtain very similar estimates as when using OLS for this restricted sample of players. Hence, I conclude that estimating the health status as a function of income, using OLS, probably gives quite accurate estimates of the causal effect. A 10 percent increase in income is thus likely to generate 0.01-0.02 standard deviations of better health for the whole population of both players and non-players. For the selected sample of players, the effect was about 2-4 times higher.

The finding that the more permanent income measure is more protective against bad health for older individuals is not found using IV-techniques. Instead, there is quite strong evidence that semi-permanent income is more protective against bad health for older people. The mortality estimates are very uncertain when lottery prizes are used as an instrument for income. There is some evidence that permanent income causally increases life expectancy. These results are not statistically significantly different from zero, however, and therefore, I do not want to stress any of the income-mortality results.

Smith (1999) suggested that using monetary lottery prizes might help us disentangle the causal pathway between income and health, which I have attempted to achieve in this study. The sample of lottery winners used in this paper is quite small, however, and many have won small prizes only, which is reflected in the standard errors when the lottery prize variable is used. Still, I found that in most estimations, the income-estimate, using only the variation in income due to monetary lottery prizes, is very similar in magnitude to the income-estimate, when using simple traditional techniques.

References

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

Björklund, Anders and Mårten Palme, "Income redistribution over the life cycle and between individuals - empirical evidence based on Swedish panel data. Forthcoming in D. Cohen, T. Piketty & G. Saint-Paul (eds.), The New Economics of Rising Inequality, Oxford University Press, 2001.

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

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

Deaton, Angus S. and Christina H. Paxson, "Aging and Inequality in Income and Health," American Economic Review, 88(2), May 1998.

Deaton, Angus S. and Christina H. Paxson, "Mortality, Education, Income, and Inequality among American Cohorts," NBER Working Paper 7140, May 1999.

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

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

Ettner, Susan L., "New Evidence on the Relationship Between Income and Health," Journal of Health Economics, 15, 1996, pp. 67-85.

Fuchs, Victor, R., "Time preference and Health: An Exploratory Study," in Economic Aspects of Health, edited by Victor R. Fuchs, pp. 93-120. Chicago: University of Chicago Press, 1982.

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

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

Grossman, Michael, "On the Concept of Health Capital and the Demand for Health," Journal of Political economy," 80, pp. 223-255, 1972.

Grossman, Michael, "The Human Capital Model for Health," Handbook of Health Economics, Eds. A.J. Culyer & J.P. Newhouse, forthcoming 2000.

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

Lindeboom, Maarten, France Portrait, and Gerard J. van den Berg, "Mental-Health Effects of Major Events in the Life of Elderly Individuals," Tinbergen Institute Disussion paper 103/3, 2001.

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

Lundberg, Olle, "Causal Explanations for Class Inequality in Health – An Empirical

Analysis," Social Science & Medicine, 32 (4), 1991, pp. 385-393.
Manderbacka, K., Questions on Survey Questions on Health, Phd Dissertation. Swedish Institute for Social Research, Stockholm, 1998.
Pritchett, Lant, and Lawrence H. Summers, "Wealthier Is Healthier," Journal of Human Resources, 31(4), pp. 841-68, Fall 1996.
Smith, James P., "Healthy Bodies and Thick Wallets: The Dual Relation Between Health and Economic Status," Journal of Economic Perspectives, 13(2), Spring 1999, pp. 145-166.
Sundberg, Gun, Essays on Health Economics, Ph.D. Dissertation, Department of Economics, Uppsala University, 1997.
Van Doorslaer, Eddy, et al., "Income related inequalities in Health: Some International Comparisons," Journal of Health Economics, 16(1), pp. 93-112, February 1997.

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.[30] 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:

$$h^* = \gamma' x + e$$

where h=2 if $\mu_2 \leq h^*$, h=1 if $\mu_1 \leq h^* \leq \mu_2$ and h=0 if $h^* \leq \mu_1$; 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 μ_j:s, like the vector γ, are left to be estimated

The 48 health symptom questions were also asked in earlier SLLS-surveys, hence the 81 indicators are also available in these waves.[31] For most of these symptoms, respondents were asked to give a statement of whether he/she had had this symptom to no, some, or a severe degree during the 12 months prior to survey. In the cases where less then 1 percent were coded in any of the two symptom categories (no problem for the first category), the some and severe group were combined.

The produced symptom-estimates are then used as weights, in that I linearly predict respondents' general health in 1981, based on the individuals' symptoms in 1981 and the ordered probit estimates, for those in the same age group, i.e. between 34-76 years of age in 1981, i.e.:

$$\hat{h}^* = \hat{\gamma}' x$$

where the estimated parameters for 1991 are shown below, in appendix table-1. Hence, $\hat{h}^*$ is constructed based on information on 48 health symptoms and their respective weights (estimated). Note that I also controlled for a cubic in age and a gender dummy when estimating the weights (which made very little difference, though).

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

[31] 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.

The predicted bad health index was then standardized to have a mean of zero and a variance of one, by dividing $\hat{h}*$ by its standard deviation

$$H = Std(\hat{h}*) \quad where \;\; H \sim N(0,1).$$

This standardized index of bad health (*H*) can then be used as a dependent variable in estimations, using linear estimation techniques (OLS, IV).

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

As shown in appendix-table 1, 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.

Appendix Table 1: Estimated health weights (using SLLS in 1991)

Health symptom	Estimate	Standard error	Fraction
DISABILITY TO MOVE			
Cannot walk 100m quite fast without probl.	.47	.10	.075
Cannot run 100m without problem	.69	.07	.236
Cannot walk in stairs without problem	.39	.09	.104
PAINS			
Pain in shoulders (some)	.21	.07	.221
(severe)	.58	.08	.129
Pain in Back/sciatica (some)	.06	.07	.228
(severe)	.45	.08	.148
Stiff/pain in the joints (some)	.15	.07	.174
(severe)	.27	.08	.114
CARDIOVASCULAR DISEASES			
Pain in chest (some)	.08	.09	.070
(severe)	.26	.14	.030
Heart weakness (some or severe)	.25	.13	.034
High blood-pressure (some)	.24	.08	.104
(severe)	.26	.15	.022
Varicose vein/ulcer (some)	.25	.09	.068
(severe)	.37	.18	.014
Swollen legs (some)	.01	.09	.077
(severe)	-.35	.16	.023
Shortness of breath (dyspnoea) (some)	.15	.11	.056
(severe)	-.17	.19	.023
Dizziness (some)	.07	.08	.093
(severe)	.10	.16	.021
MENTAL BAD HEALTH STATE			
General tiredness (some)	.35	.07	.179
(severe)	.71	.12	.050
Sleeping problem (some)	.19	.08	.108
(severe)	.31	.13	.042
Nervousness/anxiety (some)	.23	.09	.088
(severe)	.46	.17	.029
Depression (some)	.25	.12	.039
(severe)	.24	.18	.023
Mental illness (some or severe)	.60	.23	.010
STOMACH/ INTESTINAL PROBLEM			
Stomach pains (some)	-.03	.08	.138
(severe)	.27	.12	.048
Gallstone(biliary colic)/bilious(some/severe)	.07	.13	.030
Feel sick at one’s stomach (some)	-.10	.11	.080
(severe)	-.09	.20	.019
Vomiting/throwing up (some)	-.13	.15	.041
(severe)	.28	.22	.016
Diarrhea (some)	-.16	.11	.076
(severe)	-.36	.18	.019

Gastric ulcer (some or severe)	.33	.16	.023
Constipated (some)	.01	.12	.041
(severe)	.13	.19	.015
Haemorrhoids (some)	.08	.11	.051
(severe)	.30	.20	.015
OTHER			
Genital discomfort (some)	-.21	.18	.022
(severe)	.09	.22	.011
Menstrual discomfort (some)	-.03	.16	.040
(severe)	.09	.16	.030
Migraine, Headache (some)	.02	.06	.373
(severe)	.09	.09	.107
Cold, Influenza (some)	-.11	.06	.487
(severe)	-.06	.08	.165
Seeing/Eye problem not improved by glasses (some)	.33	.11	.049
(severe)	.51	.13	.031
Hearing problem (some)	.19	.07	.148
(severe)	.23	.12	.041
Chronic bronchitis/asthma (some)	.21	.12	.037
(luftvaegsinfektion) (severe)	.70	.20	.016
Struma (some or severe)	-.23	.14	.027
Heart attack, coronary (some or severe)	.24	.21	.011
Kidney problem/stone (some)	.11	.22	.012
(severe)	-.27	.22	.012
Bladder/Prostate disorder (some)	-.06	.12	.036
(severe)	.25	.17	.019
Inguinal hernia (some or severe)	.17	.18	.017
Hot flushes (sweatings) (some)	-.25	.11	.062
(severe)	-.31	.17	.022
Cough (some)	.02	.07	.182
(severe)	-.12	.13	.043
Growing thin (unnatural weight loss) (some or severe)	.19	.17	.019
Overstrained (some)	-.08	.11	.053
(severe)	-.13	.24	.010
Rashes/eczema/psoriasis (some)	.06	.08	.101
(severe)	.12	.16	.021
Cancer (some or severe)	.86	.20	.011
Anemia (blodbrist) (some or severe)	.12	.17	.022
Diabetes (some or severe)	.48	.12	.032
Overweight (some)	-.01	.08	.128
(severe)	-.12	.14	.030
Organic nervous disorder (CP,MS,Polio etc) (some or severe)	.32	.23	.009

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.

Appendix 2: OLS and Tobit regressions of health and mortality measures on log income in 1967-1981 and covariates.

Dependent variable:	Standardized index of bad health in 1981			Number of sickness symptoms in 1981			Years left alive		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log income 1967-81	-.46	-.24	-.14	-1.98	-1.03	-.64	2.44	1.63	1.45
	(.05)	(.06)	(.05)	(.27)	(.28)	(.25)	(.69)	(.75)	(.76)
Women=1	.23	.20	.10	1.78	1.57	.61	2.91	3.01	3.15
	(.03)	(.04)	(.04)	(.16)	(.18)	(.17)	(.47)	(.56)	(.56)
Age	.14	.20	.11	.55	.80	.53	-.92	-1.54	-1.38
	(.10)	(.09)	(.09)	(.49)	(.49)	(.43)	(1.70)	(1.72)	(1.72)
Age^2/k_1	-.25	-.39	-.22	-.93	-1.50	-.94	1.12	2.32	2.02
	(.20)	(.20)	(.18)	(1.02)	(1.00)	(.88)	(3.25)	(3.29)	(3.29)
Age^3/k_2	.14	.23	.13	.49	.81	.50	-.74	-1.34	-1.18
	(.11)	(11)	(.09)	(.54)	(.53)	(.46)	(1.60)	(1.62)	(1.62)
Foreign=1		.20	.10		.95	.27		1.45	1.64
		(.09)	(.08)		(.44)	(.40)		(1.04)	(1.06)
Health problems when growing up=1		.14	.07		.52	.07		-.48	-.35
		(.04)	(.04)		(.21)	(.19)		(.56)	(.56)
Economic problems when growing up=1		.20	.08		1.17	.42		-.57	-.35
		(.04)	(.04)		(.21)	(.19)		(.50)	(.51)
Years of Schooling in 1968		-.03	-.03		-.11	-.11		.02	.00
		(.01)	(.01)		(.03)	(.03)		(.09)	(.09)
Married in 1968=1		-.10	-.06		-.55	-.43		.96	.83
		(.04)	(.04)		(.21)	(.19)		(.61)	(.61)
Work in 1968=1		-.07	.02		-.53	-.15		.05	-.20
		(.05)	(.04)		(.23)	(.20)		(.65)	(.65)
Low wealth in 1968=1		.07	.05		.21	.09		-.11	-.08
		(.04)	(.03)		(.18)	(.16)		(.56)	(.56)
Very low wealth in 1968=1		.19	.03		.88	.17		-1.69	-1.34
		(.06)	(.05)		(.29)	(.24)		(.67)	(.68)
Number of sickness symptoms in 1968						.53			
						(.02)			
Standardized index of bad health in 1968			.40						-.64
			(.02)						(.23)
R^2	.217	.228	.371	.143	.174	.350			

Constant included in all estimations. Robust standard errors are shown in parentheses. Number of observations is 2948. Note that k_1=2*Mean(Age) (=106.62) and k_2=3*Mean(age^2) (=89,076.45) 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 columns 1-6, Tobit is used in columns 7-9.

Table 1: Descriptive statistics for the whole sample

Variables	Mean	St.dev	10	25	50	75	90
Main health variables							
Standardized bad Health index (STDH) in 1981	0.00	1.00	-.84	-.75	-.38	.45	1.43
Number of sickness symptoms (SYM) in 1981	5.65	4.77	1	2	4	8	12
Died before 1997 (within 15.6 years)	.24	.43					
Years left alive (YLA)	13.90	3.73	7.38	15.6	15.6	15.6	15.6
Main income variables							
Log average income 1967-81 ($\bar{I}_{81,15}$)	11.60 [11.53]	.34 [4.25]	11.18 [7.16]	11.39 [8.88]	11.61 [11.02]	11.81 [13.47]	12.00 [16.28]
Log average income 1975-81 ($\bar{I}_{81,7}$)	11.64 [12.01]	.35 [4.52]	11.20 [7.35]	11.42 [9.11]	11.65 [11.50]	11.86 [14.09]	12.05 [17.08]
Log income 1981 (I_{81})	11.63 [12.08]	.38 [6.29]	11.20 [7.34]	11.39 [8.87]	11.63 [11.26]	11.85 [14.03]	12.06 [17.20]
Player=1	.26	.44					
Demographic variables							
Age in 1981	53.31	12.26	37	42	53	64	71
Gender: women=1	.51	.50					
Family Background variables							
Foreign=1	.05	.22					
Economic problems when growing up=1	.27	.44					
Health problems when growing up=1	.21	.41					
Socioeconomic variables in 1968							
Years of schooling 1968	8.52	2.86	6	7	7	10	13
Work in 1968=1	.74	.44					
Married in 1968=1	.79	.41					
Low wealth in 1968=1	.29	.46					
Very low wealth in 1968=1	.14	.35					
Socioeconomic variables in 1974							
Years of schooling 1974	8.87	3.14	6	7	8	10	13
Work in 1974=1	.74	.44					
Married in 1974=1	.81	.39					
Low wealth in 1974=1	.21	.41					
Very low wealth in 1974=1	.10	.30					
Other covariates							
STDH in 1974	0.00	1.00	-.81	-.68	-.38	.39	1.42
SYM in 1974	5.32	4.47	1	2	4	7	12
STDH in 1968	0.00	1.00	-.78	-.62	-.37	.29	1.31
SYM in 1968	4.57	4.15	1	2	3	6	10
Log income in 1974 (I_{74})	11.55	.56	11.08	11.35	11.61	11.84	12.05
Log income in 1973 (I_{73})	11.48	.49	10.95	11.27	11.52	11.76	11.98
Log income in 1967 (I_{67})	11.28	.59	10.77	11.08	11.34	11.58	11.82

Notes: Number of observations is 2948. Log average income is the logarithm of yearly average disposable family income in adult equivalencies. For the income measures averaged over 1967-1981 and 1975-1981, 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 prizes. For details, see text (section 2). In 1998, $=SEK 9.85 according to OECD National Accounts PPP figures. Note that for three individuals, I lack observations for disposable family income in 1973. For these individuals, I replace these missing values by the average of their disposable family incomes in 1967 and 1974. STDH is the standardized index of bad health, and SYM is the number of health symptoms.

Table 2: Correlation matrix for health and mortality measures

	Standardized index of bad health in 1981	Number of sickness symptoms in 1981	Number of visits to the doctor in 1981	Years left alive
Standardized index of bad health in 1981	1			
Number of sickness symptoms in 1981	.83	1		
Number of visits to the doctor in 1981	.41	.37	1	
Years left alive	-.35	-.25	-.16	1

All p-values<0.0001 (from test of no correlation among two variables).

Table 3: OLS regressions of standardized bad health index in 1981 on log income. Less and more permanent income measures. Regressions made for all and separately by age and gender groups.

	(1)	(2)	(3)	(4)	N:
A: $\log I_{81}$					
All	-.35	-.17	-.11	-.10	2948
	(.05)	(.05)	(.04)	(.04)	
B: $\log \bar{I}_{81,7}$					
All	-.42	-.21	-.13	-.09	2948
	(.05)	(.05)	(.05)	(.05)	
C $\log \bar{I}_{81,15}$					
All	-.46	-.24	-.14		2948
	(.05)	(.06)	(.05)		
Women	-.47	-.22	-.09		1506
	(.08)	(.09)	(.08)		
Men	-.42	-.23	-.17		1442
	(.07)	(.07)	(.07)		
Age≤46	-.16	-.06	-.01		1037
	(.07)	(.07)	(.07)		
47≤Age ≤60	-.48	-.22	-.16		942
	(.10)	(.10)	(.10)		
Age≥61	-.62	-.28	-.18		969
	(.10)	(.11)	(.10)		
Age≥68	-.75	-.43	-.23		488
	(.14)	(.16)	(.14)		
D: All					
$\text{Log}\ \bar{I}_{81,15}$	-.42	-.21	-.12		2948
	(.05)	(.05)	(.05)		
$\text{Age*log}\ \bar{I}_{81,15}$	-.017	-.016	-.008		
	(.004)	(004)	(.004)		
$\text{Women*log}\ \bar{I}_{81,15}$	-.10	-.07	-.03		
	(.10)	(.10)	(.09)		
Controls:					
Women, cubic in age	Yes	Yes	Yes	Yes	
X-variables in 1968	No	Yes	Yes	No	
X-variables in 1974	No	No	No	Yes	
STDH in 1968	No	No	Yes	No	
STDH in 1974	No	No	No	Yes	

Constant included in all estimations. Robust standard errors are shown in parentheses. Number of observations is 2948. Note that the interaction terms between Age*log $\bar{I}_{81,15}$ and between Women*log $\bar{I}_{81,15}$ 7-81 are scaled so that the estimate on Log $\bar{I}_{81,15}$ y67-81 gives the average marginal effect. X-variables are the family background and socioeconomic variables listed in Table 1. STDH is the standardized index of bad health. In A, B and C; STDH in 1981 are regressed on log income in 1981, log average income in 1975-1981 and log average income in 1967-1981, respectively, and covariates. In D; STDH in 1981 is regressed on log average income in 1967-1981, and interaction terms between log average income in 1967-1981 and Age and Women, and covariates.

Table 4: Tobit regressions of years left alive (YLA) on log income. Less and more permanent income measures. Regressions made for all and separately by age and gender groups.

		(1)	(2)	(3)	(4)	Fraction uncensored observ.	N:
A: log I_{81}							
	All	2.4 (.6)	1.7 (.7)	1.6 (.7)	1.2 (.7)	.24	2948
B: log $\bar{I}_{81,7}$							
	All	2.5 (.7)	1.7 (.8)	1.6 (.8)	.9 (.8)	.24	2948
C: log $\bar{I}_{81,15}$							
	All	2.4 (.7)	1.6 (.8)	1.4 (.8)		.24	2948
	Women	2.9 (.9)	1.7 (1.1)	1.5 (1.1)		.20	1506
	Men	2.2 (1.0)	1.5 (1.1)	1.4 (1.1)		.26	1442
	Age≤46	2.9 (3.9)	5.5 (3.6)	5.5 (3.7)		.04	1037
	47≤Age ≤60	3.8 (1.6)	2.3 (1.8)	2.0 (1.8)		.16	942
	Age≥61	2.2 (.8)	1.1 (.9)	1.0 (.9)		.53	969
	Age≥68	1.7 (1.0)	1.2 (1.1)	.9 (1.1)		.66	488
D:Reg4: All						.24	
Log $\bar{I}_{81,15}$		2.6 (.9)	1.8 (.9)	1.8 (.9)			2948
Age*log $\bar{I}_{81,15}$		-.02 (.07)	-.03 (.07)	-.04 (.07)			
Women*log $\bar{I}_{81,15}$		.3 (1.3)	-.0 (1.3)	-.1 (1.3)			
Controls:							
Women, cubic in age		Yes	Yes	Yes	Yes		
X-variables in 1968		No	Yes	Yes	No		
X-variables in 1974		No	No	No	Yes		
STDH in 1968		No	No	Yes	No		
STDH in 1974		No	No	No	Yes		

Constant included in all estimations. Robust standard errors are shown in parentheses. Number of observations is 2948. Note that the interaction terms between Age*log $\bar{I}_{81,15}$ and between Women*log $\bar{I}_{81,15}$ 7-81 are scaled so that the estimate on Log $\bar{I}_{81,15}$ y67-81 gives the average marginal effect. X-variables are the family background and socioeconomic variables listed in Table 1. STDH is the standardized index of bad health. In A, B and C; STDH in 1981 are regressed on log income in 1981, log average income in 1975-1981 and log average income in 1967-1981, respectively, and covariates. In D; STDH in 1981 is regressed on log average income in 1967-1981, and interaction terms between log average income in 1967-1981 and Age and Women, and covariates.

Table 5: Descriptive statistics for players and non-players

	A: Sample of Lottery players, Player=1 (n=777)		B: Sample non-lottery play's, Player=0(n=2171)		
Variables	Mean	St.dev.	Mean	St.dev.	p-value
Main health variables					
Stand. Bad Health index (STDH) in 1981	.015	1.001	-.005	1.000	.623
Number of sickness symptoms (SYM),1981	5.59	4.69	5.66	4.81	.714
Died before 1997 (within 15.6 years)	.25	.43	.236	.425	.502
Years left alive (YLA)	13.80	3.84	13.93	3.69	.401
Main income variables					
Log average income 1967-81 ($\overline{I}_{81,15}$)	11.65	.34	11.58	.34	.000
	[12.13]	[5.02]	[11.32]	[3.91]	
Log average income 1975-81 ($\overline{I}_{81,7}$)	11.68	.34	11.62	.35	.000
	[12.55]	[5.17]	[11.82]	[4.24]	
Log income 1981 (I_{81})	11.65	.38	11.62	.38	.027
	[12.42]	[6.18]	[11.96]	[6.32]	
Main Lottery variables					
Player=1	1	0	0	0	
Average lottery prize 1969-81, in 10t, ($\overline{L}_{81,13}$)	.20	.58	0	0	
Average lottery prize 1975-81, in 10t, ($\overline{L}_{81,7}$)	.25	.99	0	0	
Average lottery prize 1969-74, in 10t, ($\overline{L}_{81t,6}$)	.15	.50	0	0	
Demographic variables					
Age in 1981	54.08	11.66	53.03	12.46	.036
Women=1	.36	.48	.57	.50	.000
Family background variables					
Foreign=1	.04	.20	.06	.23	.104
Economic problems when growing up=1	.27	.45	.27	.44	.708
Health problems when growing up=1	.22	.42	.20	.40	.325
Socioeconomic variables in 1968					
Years of schooling 1968	8.36	2.65	8.58	2.9	.054
Work in 1968=1	.82	.38	.71	.45	.000
Married in 1968=1	.76	.42	.80	.40	.048
Low wealth in 1968=1	.30	.46	.29	.45	.732
Very low wealth in 1968=1	.12	.33	.15	.36	.090
Socioeconomic variables in 1974					
Years of schooling in 1974	8.82	3.10	8.89	3.15	.562
Work in 1974=1	.81	.39	.72	.45	.000
Married in 1974=1	.79	.41	.82	.38	.018
Low wealth in 1974=1	.20	.40	.21	.41	.548
Very low wealth in 1974=1	.07	.26	.11	.31	.003
Other covariates					
STDH in 1974	.04	1.06	-.01	.98	.251
SYM in 1974	5.30	4.53	5.32	4.44	.901
STDH in 1968	-.01	.95	.00	1.02	.676
SYM in 1968	4.36	4.0	4.65	4.19	.087
Log income in 1974	11.61	.51	11.53	.57	.000
Log income in 1973	11.54	.47	11.45	.50	.000
Log income in 1967	11.31	.66	11.26	.56	.079

Notes: The number of observations is 777 in Table 1A and 2171 in Table 1B. The 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 in yearly averages in 1998 prizes. For details, see text (section 2). In 1998, $=SEK 9.85 according to OECD National Accounts PPP figures. Note that for three individuals, I lack observations for disposable family income in 1973. For these individuals, I replace these missing values by the average of their disposable family incomes in 1967 and 1974. STDH is the standardized index of bad health, and SYM is the number of health symptoms. Average lottery prizes won are expressed as yearly averages in SEK 10,000 in 1998 prizes. 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 6: Linear probability regressions of playing (player=1, non-player=0) on individual characteristics.

	(1)	(2)
Women=1	-.161	-.153
	(.016)	(.018)
Age	-.023	.003
	(.046)	(.047)
Age^2/k_1	.074	.025
	(.093)	(.095)
Age^3/k_2	-.050	-.027
	(.048)	(.049)
Foreign=1		-.054
		(.034)
Health problems when growing up=1		.024
		(.020)
Economic problems when growing up=1		-.017
		(.019)
Years of Schooling in 1968		-.005
		(.003)
Married in 1968=1		-.052
		(.021)
Work in 1968=1		.024
		(.020)
Low wealth in 1968=1		.002
		(.019)
Very low wealth in 1968=1		-.020
		(.024)
Standardized index of bad health in 1968		.013
		(.009)
Log Income in 1967		.012
		(.016)
P-value from partial F-test	---	.024
R^2	.040	.046

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-value is from a partial F-test, where the null hypotheses is that all parameters, except the ones for women and the cubic in age, are equal to zero.

Table 7: OLS/Tobit regressions of lottery prizes won (in SEK 10,000) on individual characteristics. The sample consists only of players.

Dependent variable:	Average lottery prize 1969-1981		Average lottery prize 1969-1974		Average lottery prize 1975-1981				
	OLS	Tobit	OLS	Tobit	OLS	OLS	OLS	OLS	Tobit
		Dy/dx		dy/dx					Dy/dx
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Women=1	-.11	-.14	-.12	-.23	-.10	-.05	-.05	-.05	-.15
	(.03)	(.04)	(.03)	(.10)	(.05)	(.06)	(.06)	(.06)	(.11)
Age	.14	.17	.35	.71	-.04	.03	.04	.04	-.10
	(.13)	(.15)	(.14)	(.33)	(.21)	(.23)	(.23)	(.24)	(.34)
Age^2/k_1	-.31	-.37	-.71	-1.44	.03	-.11	-.15	-.15	.11
	(.26)	(.30)	(.28)	(.66)	(.42)	(.47)	(.48)	(.48)	(.70)
Age^3/k_2	.17	.21	.36	.74	.01	.09	.11	.11	-.02
	(.13)	(.16)	(.15)	(.34)	(.22)	(.25)	(.25)	(.26)	(.37)
Foreign=1	.04	.08	.09	.14	-.01	-.03	-.03	-.03	.12
	(.06)	(.07)	(.10)	(.22)	(.06)	(.07)	(.07)	(.07)	(.18)
Health problem growing up=1	-.06	-.04	-.04	-.07	-.07	-.07	-.07	-.07	-.07
	(.03)	(.04)	(.03)	(.10)	(.05)	(.05)	(.05)	(.05)	(.10)
Economic probl. growing up=1	-.03	-.06	.01	-.00	-.05	-.06	-.06	-.06	-.14
	(.04)	(.06)	(.03)	(.09)	(.08)	(.08)	(.08)	(.08)	(.13)
S-years in 1968	-.00	-.00	.01	.03	-.015				
	(.01)	(.01)	(.01)	(.02)	(.009)				
Married in 1968=1	.01	.01	-.04	-.12	.05				
	(.05)	(.05)	(.05)	(.10)	(.07)				
Work in 1968=1	-.06	-.06	-.06	-.11	-.06				
	(.04)	(.06)	(.04)	(.12)	(.07)				
Low wealth in 1968=1	.00	.02	-.03	-.17	.03				
	(.04)	(.05)	(.04)	(.11)	(.07)				
Very low wealth in 1968=1	-.07	-.06	-.05	-.24	-.08				
	(.03)	(.05)	(.04)	(.13)	(.05)				
Standardized index of bad health in 1968	-.01	.00	.00	.02	-.02				
	(.02)	(.02)	(.02)	(.05)	(.03)				
Years of Schooling in 1974						-.02	-.02	-.02	-.03
						(.01)	(.01)	(.01)	(.02)
Married in 1974=1						-.03	-.03	-.03	-.01
						(.09)	(.09)	(.09)	(.14)
Work in 1974=1						.15	.14	.14	.16
						(.11)	(.11)	(.11)	(.17)
Low wealth in 1974=1						.03	.03	.03	.08
						(.08)	(.08)	(.08)	(.13)
Very low wealth in 1974=1						-.01	.01	.01	.08
						(.07)	(.07)	(.07)	(.15)
Standardized index of bad health in 1974						-.01	-.02	-.02	-.02
						(.02)	(.02)	(.02)	(.04)
Log income in 1974						.11	.07	.07	.07
						(.06)	(.06)	(.06)	(.12)
Log income in 1973							.08	.08	.11
							(.05)	(.05)	(.12)
Log income in 1967	.00	-.00	-.04	-.15	.04		-.01	-.01	-.04
	(.02)	(.03)	(.02)	(.07)	(.04)		(.04)	(.04)	(.09)
Average lottery prize 1969-74								.00	-.45
								(.07)	(.37)
P-value from partial F-test	.148	.615	.322	.387	.631	.765	.655	.691	.959
R^2	.017	---	.035	---	.011	.014	.015	.015	---

Constant included in all estimations. Robust standard errors are shown in parentheses. Number of observations is 777. 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 ones for women and the cubic in age, are equal to zero.

Table 8: OLS/IV regressions of the standardized index of bad health in 1981 on log income. More and less permanent income measure. The sample only consists of players.

Dependent variable:	Log income , 1967-81					Log income, 1975-81		
Sample:	All players (n=777)		Players who won 1969-1981 (n=626)			All players (n=777)		
	(1)	(2)	(3)	(4)		(5)	(6)	(7)
A:								
Log $\bar{I}_{81,k}$								
OLS	-.57	-.33	-.51	-.22		-.53	-.20	-.12
	(.10)	(.10)	(.12)	(.12)		(.10)	(.09)	(.12)
IV	-.65	-.39	-.78	-.53		-.50	-.17	-.15
	(.40)	(.38)	(.42)	(.40)		(.30)	(.26)	(.33)
B:								
Log $\bar{I}_{81,k}$								
OLS	-.54	-.30	-.52	-.23		-.53	-.21	-.16
	(.10)	(.10)	(.11)	(.12)		(.10)	(.09)	(.12)
IV	-.63	-.39	-.74	-.51		-.41	-.13	-.10
	(.36)	(.37)	(.36)	(.37)		(.23)	(.24)	(.30)
Age*log $\bar{I}_{81,k}$								
OLS	-.027	-.020	-.023	-.013		-.030	-.013	-.013
	(.007)	(.006)	(.008)	(.007)		(.007)	(.005)	(.006)
IV	-.013	-.003	-.022	-.012		-.030	-.016	-.016
	(.024)	(.022)	(.022)	(.022)		(.010)	(.009)	(.010)
Controls:					Controls:			
Women, cubic in age	Yes	Yes	Yes	Yes	Women, cubic in age	Yes	Yes	Yes
X-variables in 1968	No	Yes	No	Yes	X-variables in 1974	No	Yes	Yes
STDH in 1968	No	Yes	No	Yes	STDH in 1974	No	Yes	Yes
					Log income 67, 73 and 74	No	No	Yes
					Aver. lottery prize, 69-74	No	No	Yes

Constant included in all estimations. Robust standard errors are shown in parentheses. Note that the interaction term between Age*log $\bar{I}_{81,15}$ is scaled so that the estimate on Log $\bar{I}_{81,15}$ gives the average marginal effect. X-variables are the family background and socioeconomic variables listed in Table 5. STDH is the standardized index of bad health. In A; STDH in 1981 is regressed on log average income, respectively, and covariates. Average lottery prizes are used as the instrument for log average income. In B; STDH in 1981 is regressed on log average income 1967-1981, and interaction terms between log average income and interaction term between age and log average income, and covariates. Average lottery prizes and interaction between age and average lottery prizes are used as instruments for log average income and the interaction between age and log average income.

Table 9: Tobit regressions of years left alive(YLA) / Life expectancy on log income. More and less permanent income measure. The sample consists only of players.

Dependent variable:	Log income, 1967-81			Log income, 1975-81			
Sample:	All (n=777)	Age>60 (n=257)		All (n=777)		Age>60 (n=257)	
	(1)	(2)		(3)	(4)	(5)	(6)
Log $\bar{I}_{81,k}$							
OLS	2.57	1.58		3.14	3.98	1.94	2.02
	(1.43)	(1.61)		(1.43)	(1.57)	(1.68)	(1.82)
IV	7.60	5.63		1.15	1.50	1.09	1.31
	(6.32)	(5.13)		(2.87)	(3.31)	(2.28)	(2.47)
Controls:			Controls:				
Women, cubic in age	Yes	Yes	Women, cubic in age	Yes	Yes	Yes	Yes
X-variables in 1968	Yes	Yes	X-var. in 1974	Yes	Yes	Yes	Yes
STDH in 1968	Yes	Yes	STDH in 1974	Yes	Yes	Yes	Yes
			Log Income 67, 73, 74	No	Yes	No	Yes
			Aver. lottery prize, 69-74	No	Yes	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 5. YLA is the years left alive. YLA is regressed on log average income and covariates. Average lottery prizes are used as an instrument for log average income. STDH is the standardized index of bad health

IZA Discussion Papers

No.	Author(s)	Title	Area	Date
427	D. Del Boca	The Effect of Child Care and Part Time Opportunities on Participation and Fertility Decisions in Italy	6	02/02
428	D. Del Boca	Mothers, Fathers and Children after Divorce: The Role of Institutions	6	02/02
429	S. Anger J. Schwarze	Does Future PC Use Determine Our Wages Today? Evidence from German Panel Data	5	02/02
430	J. Schwarze M. Härpfer	Are People Inequality Averse, and Do They Prefer Redistribution by the State? Evidence From German Longitudinal Data on Life Satisfaction	3	02/02
431	M. Fertig C. M. Schmidt	The Perception of Foreigners and Jews in Germany - A Structural Analysis of a Large Opinion Survey	6	02/02
432	E. Tekin	Employment, Wages, and Alcohol Consumption in Russia: Evidence from Panel Data	4	02/02
433	J. D. Angrist A. D. Kugler	Protective or Counter-Productive? Labor Market Institutions and the Effect of Immigration on EU Natives	3	02/02
434	A. D. Kugler	From Severance Pay to Self-Insurance: Effects of Severance Payments Savings Accounts in Colombia	4	02/02
435	G. S. Epstein M. E. Ward	Perceived Income, Promotion and Incentive Effects	1	02/02
436	A. Kunze	The Evolution of the Early Career Gender Wage Gap	1	02/02
437	M. Fertig	Evaluating Immigration Policy Potentials and Limitations	6	02/02
438	A. Voicu	Employment Dynamics in the Romanian Labor Market: A Markov Chain Monte Carlo Approach	4	02/02
439	G. Fella P. Manzini M. Mariotti	Does Divorce Law Matter?	1	02/02
440	G. Bertola S. Hochguertel W. Koeniger	Dealer Pricing of Consumer Credit	7	02/02
441	C. W. Sibley P. P. Walsh	Earnings Inequality and Transition: A Regional Analysis of Poland	4	02/02
442	M. Lindahl	Estimating the Effect of Income on Health and Mortality Using Lottery Prizes as Exogenous Source of Variation in Income	3	02/02

An updated list of IZA Discussion Papers is available on the center‘s homepage *www.iza.org.*