ECuity III Project
Working Paper #10

THE EFFECT OF INCOME ON HEALTH: EVIDENCE FROM A BRITISH PSEUDO-PANEL

Matt Sutton
General Practice and Primary Care
Community-Based Sciences
University of Glasgow
m.sutton@clinmed.gla.ac.uk

Abstract

The positive correlation between income and health generates considerable social and policy concern in Britain. There is broad agreement that panel datasets are required to estimate the effect of income on health but few suitable datasets are available. This paper adopts an alternative approach based on pseudo-panels of regional British birth cohorts observed in repeated cross-sectional surveys between 1979 and 1996. Analyses of the effect of income on two measures of self-reported health are undertaken separately by gender and level of education. Both health measures are found to decline with age but to be lower in more recent birth cohorts. Increases in income are found to have a positive effect on health for all groups. In contrast to previous work, the effects of transitory income are found to be of a similar magnitude to those of permanent income. The dependence of current health on past health is found to be sensitive to the specification of unobservable heterogeneity between cohorts. The results suggest that a redistribution of income to low income groups would increase average population health and reduce income-related inequalities in health. [182 words]

January 2004

Acknowledgements
Data from the General Household Survey and the Health Survey for England were made available through the ESRC Data Archive. The copyright holder, the original data producer, the relevant funding agencies and the Data Archive bear no responsibility for their further analysis or interpretation. I am grateful to Hugh Gravelle and Andrew Jones for helpful comments on earlier versions of this paper.

Introduction

The positive correlation between income and health is well known and generates a considerable degree of research and policy concern (Department of Health, 1998; Department of Health, 2002). Relatively little is known about the mechanisms that generate this correlation, but there are three broad classes of explanation: income affects health; health affects income; and health and income are affected by common variables. Each is of social concern but warrants different policy responses. Here we concentrate on the effect of income on health and consider whether increases in income, which may be delivered by the state in the form of income support policies, are likely to result in improvements in health.

The observed correlation between income and health is likely to be a composite of the three broad types of effect. A large volume of mainly cross-sectional literature has demonstrated that the correlation is significant even after controlling for a wide range of potentially confounding factors designed to capture common-cause variation (Benzeval et al, 2001). A smaller number of cross-sectional studies exclude individuals with a history of chronic conditions, which may cause underachievement of income in the current period, in an attempt to circumvent "reverse causation" (Benzeval et al, 2001). The positive correlation between income and health reduces in size but remains significant. However, the best source of evidence is thought to be panel data studies, which enable examination of the correlation between *changes* in income and *changes* in health, conditional on health history. There are relatively few such studies but they seem to suggest that reverse-causation and selection effects are relatively minor issues (Benzeval and Judge, 2001).

There is a particular interest in the effect of income on health in Britain. Britain experiences wider income-related inequalities in health than most European countries, similar in size to the US (van Doorslaer et al, 1997) and Canada (Humphries and van Doorslaer, 2000). Britain's position in the health inequality league seems to be primarily driven by the correlation between income and health rather than the level of income inequality (van Doorslaer et al, 1997). While the size of this correlation has remained fairly constant since 1979, its magnitude is larger in Wales and Scotland than in England (Gravelle and Sutton, 2003).

Mainly due to a lack of available data, there are only three panel data studies of income and health for Britain, all of which use the British Household Panel Study (BHPS) (Benzeval and Judge, 2001; Contoyannis et al, in press; Contoyannis et al, 2003). The BHPS is an ongoing survey of individuals living in a representative set of households in 1991. It is a multi-purpose survey, examining a wide range of variables on approximately 10,000 individuals living in 5,000 households.

Benzeval and Judge (2001) use relatively simple statistical models to examine a variety of specifications for the effect of income on health in the sixth wave. They analyse four health variables, including self-assessed health and limiting longstanding illness. They measure the effect of current income on health using income quintiles, and show that the income gradient is moderated only slightly by the inclusion of initial health in the first wave. Current income, however, becomes insignificant when average income over the previous five-years is included. Further analysis reinforces the finding that average income or persistent poverty is generally more important than current income or occasional poverty. They note the important role of education and employment in determining permanent income and suggest that these should be the focus for a long run policy strategy. They also propose that future research should examine the income-health relationships for men and women separately.

Contoyannis et al (in press) estimate a dynamic model for limiting longstanding illness using seven waves of the BHPS. They focus particularly on the contributions of state dependence, heterogeneity and serial correlation in explaining health dynamics, and provide estimated effects of transitory (current) and permanent (period average) income having controlled for these factors. Their initial analysis demonstrates that separate models should be estimated for males and females. They find permanent income effects that are ten times larger than the transitory income effects. Splitting the sample by educational attainment at the end of the period, they find some evidence that, compared to those with formal qualifications, the health of individuals with no formal qualifications is more sensitive to transitory income and (for women) less sensitive to permanent income.

In a further paper, Contoyannis et al (2003) analyse self-assessed health in eight waves of the BHPS. They test for non-random attrition in the survey and find that, while there is evidence of attrition related to the variables of interest, it does not bias the coefficients on income or lagged health. They find significant differences in self-assessed health by education group only for women, and that education is more important for younger individuals. As in the previous analysis, permanent income is found to be more important than transitory income and permanent income is more important for men than women. Although the authors report no systematic pattern of income effects across education groups, there is again some evidence of greater sensitivity to transitory income and lesser sensitivity to permanent income in the groups with no formal qualifications.

Reliance solely on the BHPS for evidence of the effect of income on health has some limitations. Although Contoyannis et al (2003) have demonstrated that sample attrition does not appear to bias the coefficients of interest, the BHPS panel offers a relatively limited number of observations which gets smaller through time. Furthermore, the survey was not started until 1991. While there are now ten years follow-up available, this was a period of relatively little change in income inequality compared to the previous decade, which included the Thatcher economic reforms (Goodman and Shephard, 2002).

An alternative method for creating panel data has been proposed which involves using data from large repeated cross-sectional surveys, creating birth cohorts and analysing cohort averages as if they were observations from a true panel. The birth cohorts are created using respondents' implied years of birth and following these years of birth through repeated cross-sections. So, for example, the group of individuals aged 45 years in a 1980 survey wave are treated as follow-up observations to the group of individuals aged 44 years in the 1979 survey wave. This pseudo-panel approach has the potential advantage of avoiding some selection problems as new individuals are interviewed at each wave. It has been used successfully to examine life-cycle models of consumption and labour supply (Browning et al, 1985) and the demand for private medical insurance (Propper et al, 2003).

The General Household Survey (GHS) is a large repeated cross-sectional survey of individuals living in private households in Britain which was undertaken annually between 1971 and 1996. Prior to the inception of the BHPS it was the only all-age Britain-wide survey to include information on income and health. Consequently it has been used in several studies to examine income and health relationships (O'Donnell and Propper, 1991; van Doorslaer et al, 1997; Benzeval et al, 2001; Gravelle and Sutton, 2003). Most of these studies have used single-year waves, pooled a small number of adjacent years or treated the data as repeated cross-sections. The only study to examine cohort effects in the repeated cross-sections (Parkin et al, 1999) found significant cohort effects in the two main health variables.

This paper uses the 1979-1996 waves of the GHS to undertake a pseudo-panel analysis of income and health for individuals living in Britain. While the panel nature of the data has to be artificially created, the findings represent an alternative source of evidence on dynamic relationships between income and health for individuals in Britain to the BHPS. It is based on larger samples and covers a longer and more interesting period of observation. Moreover, it does not suffer from the potential for attrition bias caused by 'questionnaire fatigue' in long panels.

The paper is structured as follows. In the next section we describe a dynamic individual-level health function which depends on income, employment and education and time-invariant unobservable factors. We show how this can be aggregated to group (cohort) level. We describe the data and the construction of cohorts in the Data section. The results are provided in the following section. The final section discusses the findings and makes recommendations for further work.

Model

An individual-level function for health (*h*) at time *t* can be specified as:

$$h_{it} = \beta_0 + \beta_1 h_{it-1} + \beta_2 f(y_{it}) + \beta_3 U_{it} + \beta_4 g(a_{it}) + v_i + \varepsilon_{it} \quad (i = 1,...N; t = 1,...,T) \qquad [1]$$

in which y is income, a is age, $f(.)$ and $g(.)$ are (possibly non-linear) functions, U is a binary variable representing unemployment, v is an individual-specific term which does not vary over time and ε is an error term.

Equation [1] does not allow the responses of health to the variables of interest to vary across population groups and previous studies of income effects on health have identified differences by gender and by level of education. We therefore allow the β-parameters to vary across combinations of gender and educational status (denoted l; l=1, …, L).

Without true panel data it is not possible to allow for individual-specific effects. We therefore propose that individual-specific heterogeneity is largely influenced by the individual's birth cohort (j; j=1, …, J). Differences by birth cohort may reflect genetic endowments as well as preference parameters. To further allow for unobservable heterogeneity, we allow for regional variation (denoted k; k=1, …, K).

Equation [1] can then be simplified as:

$$h_{ijklt} = \beta_0 + \beta_{1l} h_{ijklt-1} + \beta_{2l} f(y_{ijklt}) + \beta_{3l} U_{ijklt} + \beta_{4l} g(a_{ijklt}) + v_{ijkl} + \varepsilon_{ijklt} \qquad [2]$$

Taking averages for birth cohort (j), region (k) and educational/gender group (l) combinations gives:

$$\bar{h}_{jklt} = \beta_0 + \beta_{1l} \bar{h}_{jklt-1} + \beta_{2l} \bar{f}(y_{jklt}) + \beta_{3l} \overline{U}_{jklt} + \beta_{4l} \bar{g}(a_{jklt}) + \bar{v}_{jkl} + \bar{\varepsilon}_{jklt} \qquad [3]$$

in which $\bar{f}(.)$ indicates the mean value of $f(y)$. This gives L data-sets of J*K groups followed up for T years. Including a one-period lag for health gives L data-sets with N=J*K*(T-1) observations.

Data

The longest-running British-wide survey that includes information on self-reported health and income is the General Household Survey (GHS). The GHS is a series of repeated cross-sections including a representative sample of individuals living in private households in England, Wales and Scotland. The GHS covers a wide range of topics during a face-to-face interview in the respondents' own homes. Information is collected from all household members. The survey was first undertaken in 1971 but the self-assessed health measure was not introduced until 1978. The public version of the 1983 wave does not contain household income (or the variables required to calculate it). In addition, the survey was not undertaken in 1997 and 1999. This analysis therefore uses data for seventeen years from 1979–1996 inclusive, excluding 1983.

We split the annual samples by gender and educational level. Educational level is a suitable stratification variable because it is generally constant for individuals beyond young adulthood. To maximise the probability of time-invariance we stratify the sample based on whether the respondent has any formal educational qualification and restrict the age-range to over sixteen years.

Although pseudo-panels do not suffer from attrition because of questionnaire fatigue or difficulties with follow-up, there remains a potential selection problem due to individuals dying or becoming ineligible because of movement into institutions (nursing homes, hospital etc.). This type of "selection effect" is of particular concern because it is health-related. We may observe, for example, cohorts appearing to become healthier in older age as the more healthy individuals survive and remain in private households. To minimise problems with this type of selection bias, and because educational information is not collected from respondents aged 70+, we restrict the analysis to individuals aged under seventy years.

The width of the birth cohorts is a trade-off between increased *N* and the accuracy of the cell averages. The combination of the 1979-1996 observation-window and the 16-69 years age restriction permits examination of individuals born between 1927 and 1963 inclusive. We exclude the last birth-year (1963) and create 9 four-year cohorts. The first cohort is born between 1927 and 1930 and ages (assuming a 1928 birth year) from 51 years to 68 years. The last cohort is born between 1959 and 1962 and ages (assuming a 1960 birth year) from 19 years to 36 years.

We also stratify the sample by region. The GHS provides twelve administrative regions. We create seven regions based on an initial analysis of the similarity of gender, age and income-standardised health: North & Yorkshire/Humberside; North West & Wales; East Midlands & West Midlands; East Anglia & South West; London; Rest of South East; and Scotland. Descriptive statistics on the number of cell observations and the number of individuals per cell are given in Table 1.

Two health variables have been collected throughout the period. The first is a measure of mainly physical functioning (Garraway et al, 1994), limiting longstanding illness (LLSI). Respondents were asked whether 'they had a longstanding illness, disability or infirmity' and, if so, whether 'it limited their activities in any way?' Combined responses to these questions are used to generate a binary measure of health status. The second is a measure of general health status, self-assessed health (SAH). Adult respondents were asked: "how do you rate your health in general?" and offered three categories of response: 'Good'; 'Fairly Good'; and 'Not Good'. In recognition of the importance of the health measures, we provide a series for each measure.

The health measures must be capable of simple aggregation. For LLSI, we take a simple proportion reporting no limiting longstanding illness. Previous studies have demonstrated the importance of including information on all categories of the SAH measure (Wagstaff and van Doorslaer, 1994; van Doorslaer and Jones, 2003). Therefore, we attach scores to each of the three SAH categories and take a mean score. The scores come from another survey, the 1996 Health Survey for England, which contains responses to the EQ-5D. Cardinal values, representing societal quality-of-life scores between death (=zero) and full health (=one), have been attributed to the respondents' EQ-5D profiles. We obtain the quantiles from the cumulative proportions of the population across the three SAH categories in the GHS, rank respondents in the HSE by their EQ-5D score, and calculate mean EQ-5D scores within each of the bands based on the GHS quantiles. The scores attached to the 'Good', 'Fairly Good' and 'Not Good' categories are 0.991, 0.775 and 0.380, respectively.

The measure of household income that has been collected throughout the period is gross household income, before tax and housing costs. Income has been converted to 1996 prices using the Retail Price Index (Office for National Statistics, 2003a). In all years household income has been equivalised using the Before Housing Costs McClements scale (Office for National Statistics, 2003b). To minimise measurement errors and reduce the influence of outliers, in each year observations below the 1st and above the 99th percentile have been excluded. We use the log of income, which is widely used and has been shown previously to approximate well the functional form for the effect of income on health (Ettner, 1996).

The unemployment measure is based on respondents' reported economic status. We calculate the ratio of the proportion "currently seeking work" to the proportion "employed", "on government schemes" and "waiting to take up a job". Individuals reporting "temporarily sick", "permanently sick", "retired", "keeping house", "students" and "other inactive" are excluded from the calculation of the cell unemployment rate.

Table 1 gives summary statistics for the pseudo-panel data. For each of the four gender by educational status groups there are 1071 observations (=9 birth cohorts *x* 7 regions *x* 17 years). There is an average of approximately 42 individual respondents within each cell for the high education groups. The average numbers of respondents within cells for the low education groups are smaller, at 19 and 26 for males and females respectively. The low education groups have lower mean logged income, higher unemployment rates and worse health for both males and females. Males have worse health on the limiting longstanding illness measure and better health using the self-assessed health score.

Analysis

There are three main parts to the analysis: (i) graphical analysis of cohort effects; (ii) simple model of age, time and regional variation; and (iii) regression models of partial effect of income on health.

Graphical analysis of cohort effects

To illustrate the cohort effects in the dataset we first aggregate the data across regions and consider smoothed time-series for each of the nine birth cohorts. The smoothed series are created using locally-weighted smoothing (lowess) with the bandwidth set at 0.8.

Simple model of age, time and regional variation

We summarise the patterns of income and health for each of the four gender by education level groups with a simple model of regional variation, the effect of ageing and time trends. Denoting the two health variables and income by x ($x = h^1, h^2, y$), we estimate the variation in mean x using a linear model including a set of regional fixed effects (α_k) and the effects of age and time using linear trends (a and t, respectively):

$$\bar{x}_{jklt} = \alpha_{0l} + \sum_{k=1}^{K-1} \alpha_{kl} D_{kl} + \delta_l a + \gamma_l t + e_{jklt}$$

in which e is a zero-mean error term. For ease of presentation of the coefficients, we express age and time as differences from their minimum values (19 and 1979, respectively) and divide both variables by 100. The constant term is therefore the mean value for the most recent birth cohort in the base region in 1979. We measure the extent of regional variation using the standard deviation of the α_k-coefficients (including the base region value of zero).

Regression model of partial effect of income on health

Both of the health variables are limited to the {0,1} range. We therefore use a log-odds transformation for both measures: $\bar{h}^* = \ln\left(\frac{\bar{h}}{1-\bar{h}}\right)$. The transformed variable, $\bar{h}^*$, is undefined when $\bar{h}$=0 or $\bar{h}$=1. As this happens rarely and for cells with few observations, these missing observations are excluded. Since the income variable is also log-transformed, the income coefficient represents the partial elasticity of (a) the odds of no LLSI and (b) the odds of perfect SAH with respect to income.

Equation [3] specifies a lagged effect of past health status on current health and so we drop the periods for which there are no data for the preceding year (1979 and 1984) from the initial analysis. Equation [3] also specifies a cohort effect, captured by the $\bar{v}_{jkl}$ term. To allow for correlation between the cohort effects and the time-variant variables, we estimate the cohort effect as a series of dummy variables specific to each regional birth cohort for each gender/educational status group (u_{jkl}). Previous research has also shown different effects of 'permanent' and 'transitory' income on health (Contoyiannis et al, 2003). 'Permanent' income is often specified as the mean level of income for the individual (in this case, cohort) across the entire time period. The mean level of income for each cohort does not vary over time and so is perfectly correlated with u_{jkl}. We therefore first estimate a simpler model in which the cohort effects are specific to each birth-cohort and gender/educational group (u_{jl}) and create a measure of permanent income by taking the period average of income for each of the *j-k-l* combinations ($\bar{f}(y_{jkl})$).

To allow for the variation in cell sizes we estimate the two specifications of equation [3] using weighted least squares estimates of the grouped logit model. The weights are calculated using (Statacorp, 2003):

$$w_{jklt} = \frac{N_{jklt}\bar{h}_{jklt}(1-\bar{h}_{jklt})}{\bar{h}_{jklt} + (1-\bar{h}_{jklt})} \qquad [4]$$

in which N_{jklt} is the number of individual respondents on which the cell averages are based. We allow for additional heteroskedasticity across the *j-k-l* combinations using robust standard errors clustered by cohort.

Results

Graphical analysis of cohort effects

Figure 1 shows the series for the absence of limiting longstanding illness with the figures separated by gender and educational status. The top left panel shows the series for males with some formal educational qualifications. The upper left series in this panel represents the most recent birth cohort, born between 1959 and 1962 and ageing from 19 years to 36 years (assuming a 1960 birth-year). As expected the proportion without LLSI declines as the cohort ages. At age 23 years, these data are joined by a second series representing the second most recent birth-cohort, born between 1955 and 1958 and ageing from 23 years to 40 years (assuming a 1956 birth-year). The most recent birth-cohort has a lower probability of being without LLSI than the second most recent birth-cohort for the most of the age-range over which their observations overlap. This is the general trend that is observed – each cohort tends to become less healthy as it ages and, where the observations on cohorts overlap, the more recent birth cohorts tend to have lower probabilities of being without LLSI. This suggests a cohort-related decline in health as well as the familiar age-related decline in health.

The bottom-left panel of Figure 1 shows a similar set of figures for females with some formal educational qualifications. The series show a similar cohort-related decline in health, but the effect of ageing is generally weaker for educated females. At the age of 60-65 at the end of the period, educated females have a 10% higher probability of being without LLSI than educated males. The two panels on the right-hand side of Figure 1 show the series for males and females with no formal qualifications. For both genders, individuals with and without formal qualifications tend to have similar probabilities of being without LLSI in early adulthood but health declines more rapidly with age for those with no qualifications. Females with no qualifications, for example, reach the same probability of being without LLSI in their early-50s as females with qualifications reach in their late-60s.

Figure 2 shows a similar set of profiles using the mean self-assessed health score as the health measure. Here the cohort effects are also evident but appear somewhat weaker. This cohort-related decline appears for all four gender and educational status groups, although the cohort differences seem larger for the low education groups. For both males and females, the low education group appears to have a lower level of health than the high education group at the youngest age, and their health profile tends to decline more quickly with advancing age.

Figure 3 shows smoothed time-series for real-terms, equivalised gross annual household income. The differences by educational attainment are substantial. For individuals with some qualifications, the time trends within cohorts are strong and positive up to the age of 55. The general picture is one of rising incomes between cohorts. Where the cohort observations overlap, the more recent birth-cohorts tend to have higher incomes. The patterns of income change within low education groups are similar to those observed in the high education groups but within much narrower ranges.

Simple model of age, time and regional effects

Table 2 presents results of the simple descriptive analysis of the age and time trends and regional variations in the self-assessed health score and equivalised income. The results for the absence of LLSI are similar to those for self-assessed health and are not shown.

Comparison of the values of the constant terms across education groups confirms that the low education groups begin the period with lower levels of self-assessed health at young ages. The age-related declines in health are strongest for the low education groups and for males. There are also trend declines in health, which are largest for males and low education groups. Health in the base region (the North East of England, Yorkshire and Humberside) is lowest for educated males. For educated females, health is lowest in London. For the low education groups, health is lowest in the North West and Wales. The regional variation in health is largest for males and the low education groups.

The descriptive analysis of income shows that incomes are higher for the educated groups. The time trends are strongest for the educated groups. For educated males, incomes have risen at an average rate of £400 per annum and, for educated females, at £330 per annum. The annual trend is approximately one-fifth of the size for low education groups but the age-related increases tend to be larger. Incomes are lowest in the base region for the educated groups and lowest in Scotland for the low education groups. Incomes are highest in the South East of England, including London. Regional variations in income are largest for the educated groups and for males.

Regression models including permanent income

In all cases the estimated effects of unemployment are not significantly different from zero and have been omitted from the models. The results for the self-assessed health score for the four gender by education status groups are shown in Table 3. In all of the models the cohort effects are significant and show a trend decline in health – *ceteris paribus*, more recent birth-cohorts experience a lower average health score. These cohort effects are larger for males and for the two high education groups.

The estimated elasticities for transitory income are slightly less than 0.5 for all groups, except females with educational qualifications for whom the estimate is 0.3. The effect of permanent income is significant and positive for three of the four groups and tends to be larger for the groups without formal qualifications. The effect of permanent income is larger than the effect of transitory income only for males with no formal qualifications. Lagged health has a positive effect on current health for all groups but is of marginal significance. A substantially higher proportion of the variation in health is explained by these factors for males compared to females.

Table 4 presents the results for self-assessed health when sixty-three fixed effects are introduced for combinations of region and birth-cohort. The effects of lagged health become insignificant and have been omitted so that the analysis can encompass the data for all seventeen years. The elasticities of health with respect to transitory income are similar to those in Table 3 and remain positive and significant for all four groups.

These patterns of results are broadly replicated when the absence of LLSI is used as the health measure (Tables 5 and 6). In Table 5 the income-elasticity estimates are slightly larger for LLSI than self-assessed health but the pattern across groups is similar. Both male groups and females without formal qualifications show a similar effect of transitory income. Permanent income is more important for the groups without formal qualifications. The effect of income is smallest for females with formal qualifications. As for self-assessed health, there is little evidence of an effect of lagged health on current health. When the full set of fixed effects are included

(Table 6), the elasticities of health with respect to transitory income remain positive and significant for all groups. Only the coefficient for males with no formal qualifications shows any material change in magnitude.

Discussion

This paper has adopted a pseudo-panel approach to estimating the effect of income on health by following cohorts defined by gender, education, birth-cohort and region of residence through a seventeen-year period. Health is measured in two different ways, using variables that were included in the 2001 decennial Census of the UK population. The reliance on self-reported health measures could be seen as a limitation of the present study but these measures have been shown to predict 'harder' health end-points such as mortality (Idler and Benyamini, 1997) and to be comparable across social groups (van Doorslaer and Gerdtham, 2003).

The main findings of this analysis are broadly in line with previous studies using British data. Transitory and permanent income have significant effects on health, even when we have controlled for a range of confounding factors including gender, age, region of residence and educational attainment. There are differences however in the relative magnitudes of the transitory and permanent income effects and the patterns across population groups.

Previous studies have indicated that the permanent income effects are substantially larger than the transitory income effects (Contoyannis et al, 2003; Contoyannis et al, in press). In this paper, the effects are more similar in magnitude. This is likely to reflect the different units of analysis. At the individual-level, the larger effects of permanent income measures may reflect greater measurement error in the transitory income variable. At the cohort-level, there is less independent variability in the income measures with which to estimate their separate effects.

Contoyannis et al's analyses also seem to indicate that transitory income is more important, and permanent income less important, for low education groups. In this paper, the opposite results are found with the permanent income elasticities larger for the low education groups. In common with Contoyannis et al's analyses, however, permanent income is found to be more important for men than women.

A further difference in the results of this paper is the effect of health history. While previous papers based on individual-level data find significant state dependence, measures of lagged health for cohorts in the paper are of marginal significance and become insignificant when we allow for regional heterogeneity. Again, this is most likely a function of the units of analysis. At the individual-level there is significant history in the health data, while at the group-level historical differences in health between cohorts are unimportant in predicting current health.

Nevertheless, the observed differences in health between birth cohorts are significant and large. They have implications for the interpretation of cross-sectional differences in health between age groups. These results suggest that health declines more rapidly with age than cross-sectional profiles imply. The cross-sectional profiles are confounded by cohort effects with more recent birth cohorts having lower age-adjusted health than earlier cohorts. This trend also confounds the simple correlation between income and health over time, so that repeated cross-sectional surveys of the population demonstrate rising average incomes with no corresponding improvements in average health. As the health measures are self-reported it is unclear whether the decline in health between birth-cohorts reflect objective differences in health or rising expectations. Either is likely to have important consequences for future health care consumption and social policy.

The models estimated in this paper include few independent variables, although the included variables represent the main effects on health: gender, age, income, education, employment, region of residence and year-of-birth. The finding of an insignificant effect of unemployment is surprising and investigation of further influences could form an important component of future work. It would also be useful to investigate alternative stratifications for the definition of cohorts. While region of residence appears to be a significant stratification variable it is not strictly time-invariant as individuals may migrate to high-income areas and present a selection problem. Finally, this study has investigated the combined effect of all income changes on health. Future work could concentrate on whether income increases have the same effect as income falls and whether different sources of income matter (Benzeval and Judge, 2001). It would also be fruitful to evaluate the actual effects of previous changes in the welfare and tax system on the health of different groups.

The concave relationship between income and health that has been observed in cross-sectional studies suggests that redistributing income from rich to poor would increase mean population health (Gravelle, 1998). The finding of a positive effect of logged income on health for all population groups in this pseudo-panel adds further evidence to support policies designed to improve population health through income support.

[5,155 words]

References

Benzeval M, Judge K, Shouls S. Understanding the relationship between income and health: how much can be gleaned from cross-sectional data? *Social Policy and Administration* 2001; 35(4): 376-96.

Benzeval M, Judge K. Income and health: the time dimension. *Social Science and Medicine* 2001; 52: 1371-1390.

Cohen G, Forbes J, Garraway M. Interpreting self reported limiting long term illness. *BMJ* 1995; 311: 722-724.

Contoyannis P, Jones AM, Rice N. Simulation-based inference in dynamic panel probit models: an application to health. *Empirical Economics* (in press).

Contoyannis P, Jones AM, Rice N. *The dynamics of health in the British Household Panel Survey*. ECuity III Project Working Paper #3. August 2003. [Forthcoming in *Journal of Applied Econometrics*].

Department of Health. *Independent Inquiry into Inequalities in Health Report.* London: The Stationery Office 1998.
[http://www.archive.official-documents.co.uk/document/doh/ih/ih.htm]

Department of Health. *Tackling Inequalities in Health: 2002 Cross-Cutting Review*. London: The Stationery Office, 2002.
[http://www.doh.gov.uk/healthinequalities/ccsrsummaryreport.htm]

Ettner S. New evidence on the relationship between income and health. *Journal of Health Economics* 1996; 15: 67-85.

Goodman A, Shephard A. *Inequality and living standards in Great Britain: some facts*. Briefing Note 19. Institute for Fiscal Studies: London, 2002.

Gravelle H. How much of the relation between population mortality and unequal distribution of income is a statistical artefact? *British Medical Journal* 1998; 316: 382-5.

Gravelle H, Sutton M. Income related inequalities in self assessed health in Britain: 1979–1995. *Journal of Epidemiology and Community Health* 2003; 57: 125-129.

Humphries KH, van Doorslaer E. Income-related health inequality in Canada. *Social Science & Medicine* 2000; 50(5): 663-671.

Idler EL, Benyamini Y. Self-rated health and mortality: a review of twenty-seven community studies. *Journal of Health and Social Behaviour* 1997; 38: 21-37.

O'Donnell O. and Propper C. Equity and the distribution of UK National Health Service resources. *J Health Econ* 1991; 10: 1-19.

Office for National Statistics. *Retail Prices Index: Index numbers of retail prices 1948-2003*. [http://www.statistics.gov.uk/rpi]

Office for National Statistics. *McClements equivalence scale (before housing costs)*. http://www.statistics.gov.uk/Harmony/Secondary/income.asp

Parkin D, Rice N, Sutton M. Non- and semi-parametric estimation of age and time heterogeneity in repeated cross-sections: an application to self-reported morbidity and general practitioner utilization. *Health Economics* 1999; 8: 429-440.

Statacorp. *Methods and formulas section for glogit command*. Reference Manual State version 8. Stata 2003.

Van Doorslaer E, Gerdtham U-G. Does inequality in self-assessed health predict inequality in survival by income? Evidence from Swedish data. *Social Science & Medicine* 2003; 57(9): 1621-9.

van Doorslaer E, Jones AM. Inequalities in self-reported health: validation of a new approach to measurement. *Journal of Health Economics* 2003; 22(1): 61-87.

van Doorslaer E, Wagstaff A, Bleichrodt H, Calongec S, Gerdtham U-G, Gerfin M, Geurts J, Gross L, Häkkinen U, Leu RE, O'Donnell O, Propper C, Puffer F, Rodríguez M, Sundberg G, Winkelhake O. Income-related inequalities in health: some international comparisons. *Journal of Health Economics* 1997; 16(1): 93-112.

Table 1: Descriptive statistics

Variable	Male Some formal qualifications	Male No formal qualifications	Female Some formal qualifications	Female No formal qualifications
Cell observations	1,071	1,071	1,071	1,071
Individual observations	45,273	25,101	41,554	33,644
Individuals within cells				
Mean	42.3	23.4	38.8	31.4
Minimum	6	3	5	4
10th percentile	20	11	15	15
25th percentile	27	15	21	21
Median	39	22	34	30
75th percentile	55	30	52	40
90th percentile	70	38	70	50
Maximum	119	72	131	84
Variables				
Mean Log(income)	9.727	9.363	9.694	9.315
Mean age (years)	40.182	44.469	39.180	44.777
Unemployment rate	0.056	0.133	0.044	0.062
Proportion without LLSI	0.854	0.762	0.854	0.768
Mean self-assessed health score	0.913	0.852	0.896	0.829

Table 2: Descriptive analysis of age, time and regional effects

Gender Education group	Male Some formal qualifications	Male No formal qualifications	Female Some formal qualifications	Female No formal qualifications
Self-assessed health				
Constant term	0.9574	0.9475	0.9237	0.8876
Age[a]	-0.2432	-0.3780	-0.1456	-0.2514
Time[b]	-0.0591	-0.1867	-0.0439	-0.0591
N. West & Wales	0.0046	-0.0080	-0.0007	-0.0047
E.Mids. & W.Mids.	0.0143	0.0125	0.0053	0.0082
E.Anglia & S.East	0.0210	0.0420	0.0156	0.0386
London	0.0130	0.0185	-0.0026	0.0004
Rest of S.East	0.0241	0.0456	0.0116	0.0367
Scotland	0.0041	0.0043	0.0073	0.0008
N.East & Yorks/Humb[c]	0.0000	0.0000	0.0000	0.0000
Std.Dev.	*0.0091*	*0.0206*	*0.0068*	*0.0183*
Income				
Constant term	13.877	10.781	13.235	10.236
Age	1.276	4.088	6.914	5.412
Time	40.635	8.038	32.696	5.004
N. West & Wales	0.452	0.131	0.112	0.052
E.Mids. & W.Mids.	1.175	0.677	0.755	0.509
E.Anglia & S.East	1.238	1.014	0.704	1.103
London	5.257	3.306	4.237	2.670
Rest of S.East	4.338	2.979	3.564	2.786
Scotland	0.264	-0.221	0.467	-0.326
N.East & Yorks/Humb[c]	0.000	0.000	0.000	0.000
Std.Dev.	*2.102*	*1.442*	*1.738*	*1.283*

Notes

[a] Age measured as *(age-19)/100.*

[b] Time measured as *(year-1979)/100.*

[c] Base region.

Table 3: Regression of self-assessed health score with birth-cohort effects

Gender	Males		Males		Females		Females	
Education	Some formal qualifications		No formal qualifications		Some formal qualifications		No formal qualifications	
Variable	Coeff.	t-ratio	Coeff.	t-ratio	Coeff.	t-ratio	Coeff.	t-ratio
Transitory income	0.459	5.7	0.487	5.6	0.301	3.7	0.477	5.6
Permanent income	0.356	2.3	0.788	4.0	-0.072	-0.5	0.489	2.6
Lagged health	0.044	1.5	0.053	1.8	0.039	1.3	0.071	1.9
Age	-24.765	-2.3	-12.371	-0.7	15.746	1.1	36.840	2.5
Age^2	99.590	2.4	65.816	1.0	-49.489	-0.9	-120.764	-2.4
Age^3	-192.231	-2.9	-155.156	-1.6	51.867	0.6	151.093	2.1
Age^4	126.626	3.3	113.837	2.1	-16.094	-0.3	-63.720	-1.6
Born 1931-1934	0.006	0.1	-0.105	-1.3	-0.027	-0.4	-0.049	-1.0
Born 1935-1938	-0.135	-1.8	-0.177	-2.2	0.037	0.5	-0.095	-1.6
Born 1939-1942	-0.207	-2.3	-0.218	-2.5	-0.116	-1.4	-0.108	-2.0
Born 1943-1946	-0.258	-4.0	-0.280	-3.3	-0.124	-1.6	-0.149	-3.0
Born 1947-1950	-0.307	-4.3	-0.327	-3.9	-0.093	-1.2	-0.089	-1.4
Born 1951-1954	-0.417	-5.2	-0.449	-4.4	-0.174	-2.0	-0.111	-2.0
Born 1955-1958	-0.542	-6.8	-0.459	-4.7	-0.234	-2.7	-0.119	-2.0
Born 1959-1962	-0.605	-7.1	-0.484	-4.9	-0.266	-3.0	-0.108	-1.5
Constant term	-2.210	-1.2	-8.307	-3.1	-1.320	-0.6	-10.798	-5.2
R^2	0.617		0.577		0.298		0.394	
N	945		945		945		945	

Table 4: Regression of self-assessed health score with region & birth cohort effects

Gender	Males		Males		Females		Females	
Education	Some formal qualifications		No formal qualifications		Some formal qualifications		No formal qualifications	
Variable	Coeff.	t-ratio	Coeff.	t-ratio	Coeff.	t-ratio	Coeff.	t-ratio
Transitory income	0.453	5.6	0.440	5.6	0.304	3.6	0.420	5.8
Age	-21.917	-2.3	-16.736	-1.0	7.691	0.6	20.679	1.6
Age^2	92.008	2.5	77.347	1.3	-19.449	-0.4	-67.983	-1.5
Age^3	-183.146	-3.1	-168.527	-1.8	2.298	0.0	76.332	1.1
Age^4	122.436	3.5	119.115	2.3	13.642	0.3	-25.360	-0.7
Constant term	0.710	0.7	0.025	0.0	-1.262	-0.9	-3.843	-2.3
Adj-R^2	0.621		0.622		0.335		0.450	
N	1,071		1,071		1,071		1,071	

Regression model includes 63 dummy variables for birth cohort by region effects (not shown).

Table 5: Regression of absence of limiting longstanding illness with birth-cohort effects

Gender	Males		Males		Females		Females	
Education	Some formal qualifications		No formal qualifications		Some formal qualifications		No formal qualifications	
Variable	Coeff.	t-ratio	Coeff.	t-ratio	Coeff.	t-ratio	Coeff.	t-ratio
Transitory income	0.616	5.7	0.551	4.3	0.335	2.5	0.627	4.3
Permanent income	0.420	2.4	1.010	3.6	-0.159	-0.6	0.744	2.6
Lagged health	0.031	1.3	0.066	2.0	-0.008	-0.3	0.024	0.9
Age	-52.457	-2.5	-52.681	-2.0	9.331	0.5	4.974	0.2
Age^2	197.131	2.6	198.835	2.1	-42.482	-0.6	-11.424	-0.1
Age^3	-344.787	-2.9	-349.835	-2.4	45.807	0.4	-29.057	-0.2
Age^4	211.808	3.1	215.861	2.7	-13.538	-0.2	43.702	0.6
Born 1931-1934	-0.063	-0.5	-0.190	-1.8	-0.011	-0.1	-0.199	-2.7
Born 1935-1938	-0.165	-1.4	-0.338	-3.2	-0.060	-0.6	-0.283	-4.1
Born 1939-1942	-0.280	-2.5	-0.389	-4.0	-0.126	-1.2	-0.415	-4.5
Born 1943-1946	-0.326	-2.7	-0.430	-3.9	-0.233	-2.1	-0.434	-5.5
Born 1947-1950	-0.432	-3.3	-0.578	-5.1	-0.210	-1.9	-0.508	-5.0
Born 1951-1954	-0.508	-3.9	-0.652	-4.7	-0.308	-2.7	-0.503	-5.0
Born 1955-1958	-0.634	-4.7	-0.820	-5.8	-0.296	-2.3	-0.563	-5.4
Born 1959-1962	-0.727	-5.0	-0.850	-5.7	-0.482	-3.7	-0.643	-5.5
Constant term	-1.763	-0.6	-6.623	-1.7	0.753	0.3	-10.386	-2.9
R^2	0.597		0.553		0.498		0.518	
N	929		870		922		899	

Table 6: Regression of absence of limiting longstanding illness with region & birth cohort effects

Gender	Males		Males		Females		Females	
Education	Some formal qualifications		No formal qualifications		Some formal qualifications		No formal qualifications	
Variable	Coeff.	t-ratio	Coeff.	t-ratio	Coeff.	t-ratio	Coeff.	t-ratio
Transitory income	0.601	5.5	0.484	3.7	0.326	2.5	0.634	4.8
Age	-41.162	-2.5	-64.762	-2.5	20.504	1.1	5.036	0.2
Age^2	156.566	2.4	232.579	2.4	-85.841	-1.2	-12.524	-0.1
Age^3	-283.894	-2.7	-390.143	-2.7	115.061	1.0	-28.024	-0.2
Age^4	178.577	2.9	232.463	2.9	-52.996	-0.8	43.280	0.6
Constant term	1.102	0.6	4.786	1.6	-1.894	-1.0	-3.665	-1.1
Adj-R^2	0.600		0.573		0.512		0.552	
N	1,061		1,024		1,056		1,043	

Regression model includes 63 dummy variables for birth cohort by region effects (not shown).

Figure 1: Age and cohort effects on absence of limiting longstanding illness

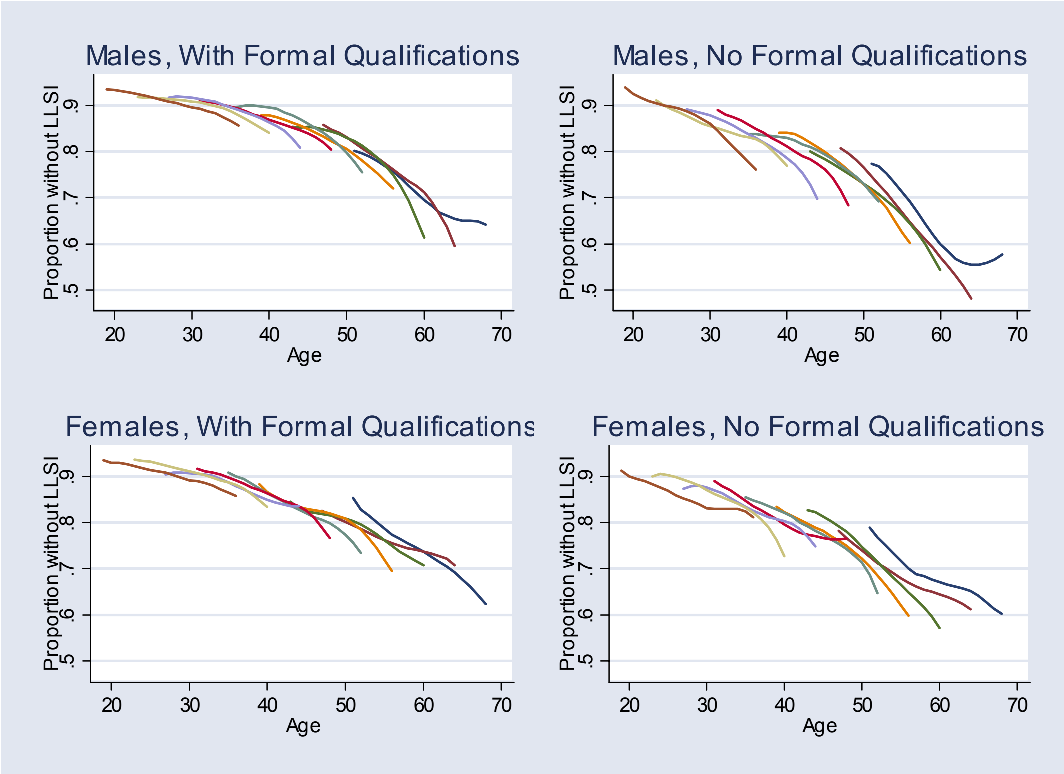


Figure 2: Age and cohort effects on self-assessed health score

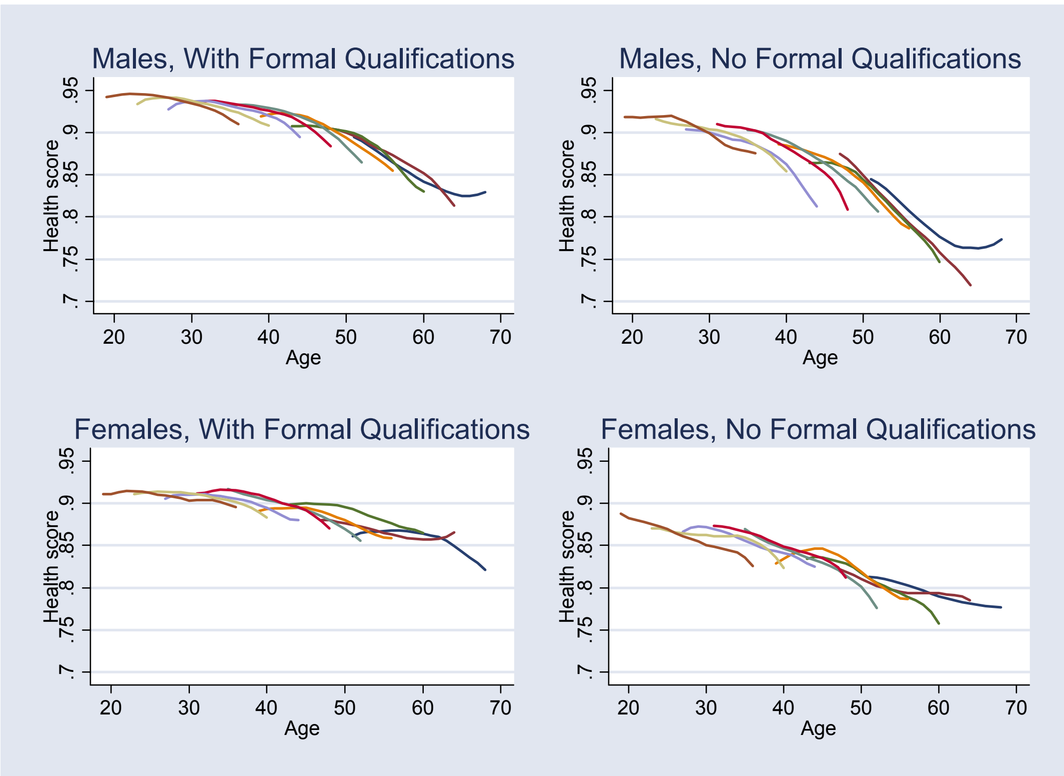
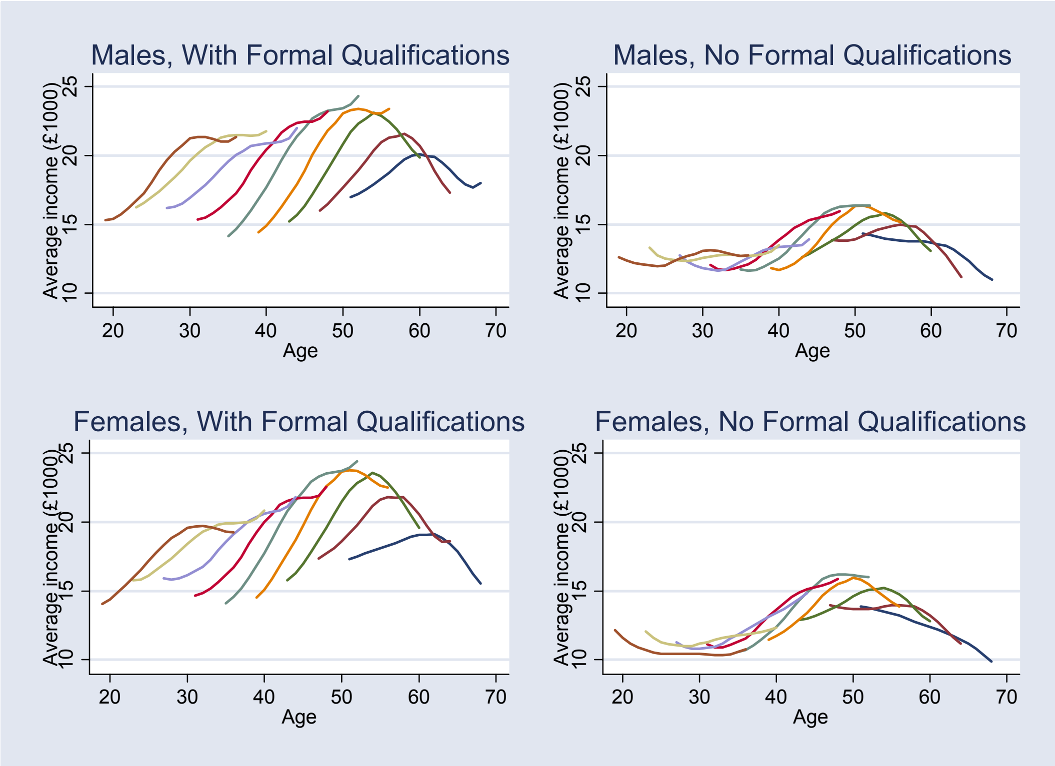


Figure 3: Age and cohort effects on real-terms equivalised gross household income (1996 prices)[1]



[1] Figures are equivalised to convert observed income to an equivalent level of income for a household containing two adults and no children.