J. R. Statist. Soc. A (2006)
169, *Part* 2, *pp.* 337–359

Pathways between education and health: a causal modelling approach

Tarani Chandola,

University College London, UK

Paul Clarke and J. N. Morris

London School of Hygiene and Tropical Medicine, UK

and David Blane

Imperial College London, UK

[Received October 2004. Final revision October 2005]

Summary. The association of poor education and poor health has been consistently observed in many studies and in various countries. Thus far, studies examining the mechanisms underlying this association have looked at only a limited set of potential pathways. This study simultaneously examines six distinctive pathways, which have been hypothesized to link education and health and found support from previous studies. A causal analysis of education and health was performed using structural equation models. Data were used from six phases of the National Child Development Study, which is based on following up an initial sample of 17 416 children who were born in 1958. The association between education and health appears to be explained by a combination of mechanisms: adolescent health and adult health behaviours for men and women, adult social class among men and parental social class among women. We conclude that improvements in population educational attainment may not automatically lead to improvements in population health, and that health policies for improving health and reducing health inequalities need to target specific causal pathways.

Keywords: Education; Health; Pathways; Structural equation modelling

1. Introduction

In 1992, the UK Department of Education first published league tables of school leaving pass rates in the different local education authority areas of England. The response of one experienced epidemiologist (JNM) on viewing the league tables was 'this list looks familiar', because it closely resembled a list of the same areas ranked by mortality. An analysis of these results by Morris *et al*. (1996) confirmed this impression: at the local education authority level, there was a strong correlation between the all-cause mortality rate and the proportion of children who are aged 15–16 years with passing grades.

Blane *et al.* (1996) looked further at the relationship between pass rate, mortality and deprivation in the local education authority areas. They found that deprivation explained much of the relationship between education and mortality, whereas little of the relationship between

Address for correspondence: Tarani Chandola, International Institute for Society and Health, Department of Epidemiology and Public Health, University College London, 1–19 Torrington Place, London, WC1E 6BT, UK.
E-mail: t.chandola@ucl.ac.uk

 0964–1998/06/169337

mortality and deprivation was explained by education. A naïve interpretation of this finding would be that deprivation affects mortality, and that deprivation also plays some mediating role for the relationship between education and mortality. In reality, as Blane *et al.* (1996) acknowledged, the conclusions that can be drawn from this analysis are extremely limited. Not only does the ecological fallacy preclude drawing reliable inferences about individuals from aggregated data, but also any tentative causal interpretation of the results is prevented by the fact that pass rates measure educational attainment of adolescents, whereas mortality rates quantify deaths among (mainly) older individuals. However, the above illustration demonstrates the need to understand and quantify the effect of the pathways linking education and health and how these could contribute to the debate on the role of education in reducing health inequalities.

Studies of education and health have already been undertaken using disaggregated, individual level data, particularly in economics (e.g. Berger and Leigh (1989), Feinstein (2002) and Murasko (2003)), epidemiology (e.g. Davey Smith *et al*. (1998) and White *et al*. (1999)) and sociology (e.g. Ross and Wu (1995)). More comprehensive reviews of the area can be found elsewhere (Blane, 2003; Mirowsky and Ross, 2003). In each of these studies, statistical models were used to estimate the direct effect of education on health, with adjustment for factors that could induce a spurious association between the two. However, only one study to date has explicitly modelled multiple pathways in its analysis of education and health (Mirowsky and Ross, 1998).

2. Pathways between education and health

Blane (2003) reviewed the literature in which the links between education and health were investigated and highlighted five plausible pathways for which evidence has been consistently found. In addition, we have added cognitive ability to this list (Feinstein, 2002). These six pathways are illustrated in Fig. 1.

(a) *Cognitive ability* (or intelligence) in childhood may confound the association between education and health as it affects both educational attainment and adult health outcomes (Gottfredson, 2004; Gottfredson and Deary, 2004; Batty and Deary, 2004). Cognitive ability in learning, reasoning and solving problems affects academic performance positively as well as job performance and functional literacy (Gottfredson, 2004). Furthermore, intelligence predicts lower mortality rates from all causes, cardiovascular disease and lung cancer (Hart *et al*., 2003). Intelligence could enhance a person's care of their own health through better receptivity to health education messages, by representing learning, reasoning and problem solving skills that are useful in preventing chronic disease and accidental injury and in completing complex treatment regimes (Gottfredson and Deary, 2004).

(b) *Childhood socio-economic circumstances* may confound the education–health association as it influences educational attainment as well as health. Higher parental social class is strongly associated with greater parental interest in the child's education and, consequently, better educational attainment (Feinstein and Symons, 1999). Lower parental social class (in childhood) has also been strongly associated with greater morbidity in adulthood (Power and Hertzman, 1997). The association between education and health could therefore arise from the effect of parental social class on educational attainment and adult health.

(c) *Childhood and adolescent health* may confound the association between education and health as it could influence a child's educational attainment (Isohanni *et al*., 2001) and

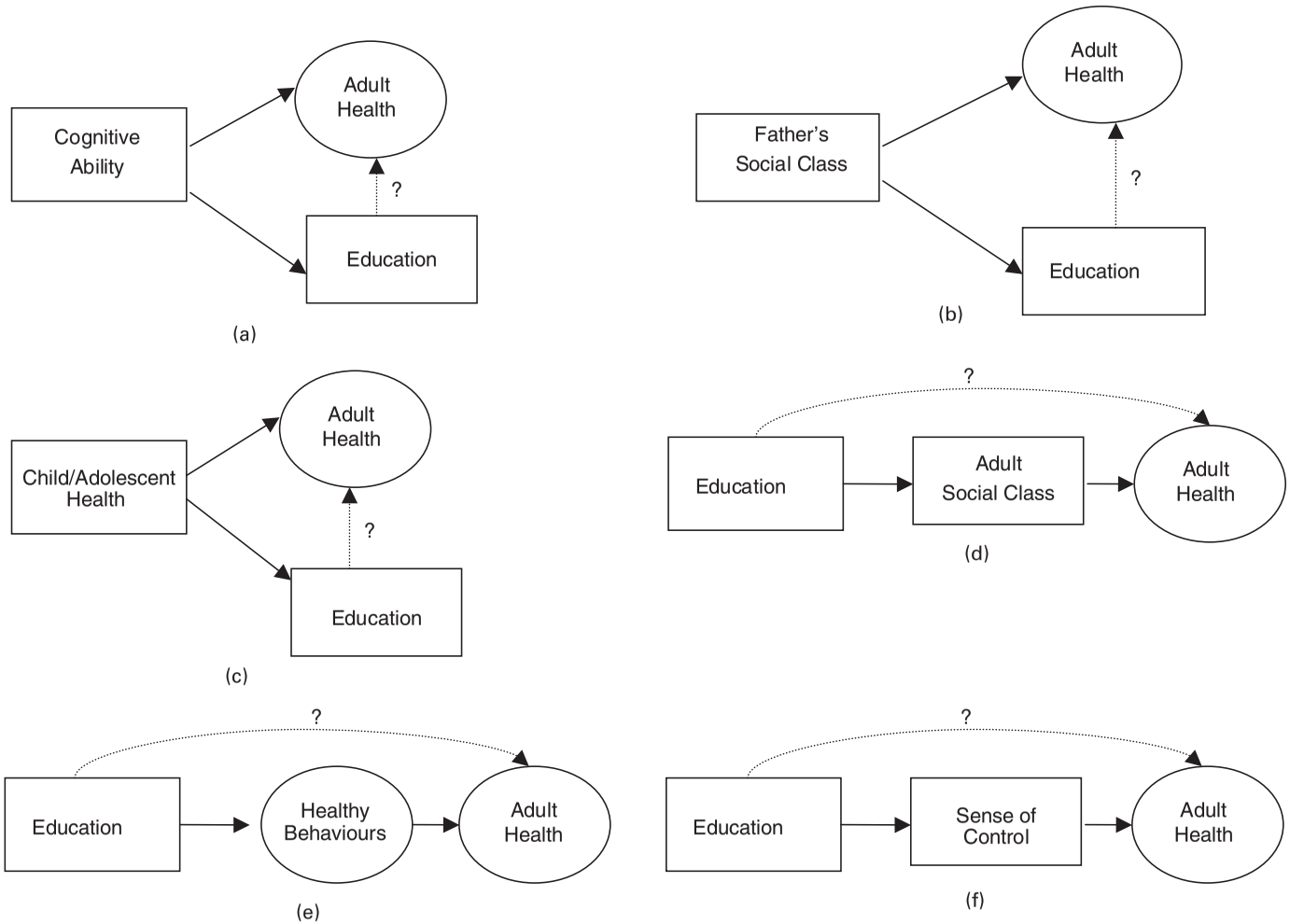

Fig. 1. Pathways between education and health: (a) cognitive ability; (b) father's social calss; (c) child and adolescent health; (d) adult social class; (e) healthy behaviours; (f) sense of control

their later, adult, health (Jefferis *et al*., 2002). Ill-health or disease during childhood limits educational attainment and predisposes individuals towards adult morbidity (Hotopf *et al*., 1999).

(d) *Adult socio-economic circumstances* may mediate the effect of educational attainment on adult health. Some studies have found that the effect of education on adult health becomes non-significant after adjusting for occupational class and income (Dahl, 1994; Davey Smith *et al*., 1998). This suggests that the effect of education on health operates through a pathway where higher education leads to higher occupational attainment and income, which, in turn, increases the chances of better health. In contrast, other studies have found that the effect of education on health remains strong even after controlling for occupational class and income (Winkleby *et al*., 1992).

(e) *Health behaviours* may mediate the effect of education on health as education might affect a person's receptivity to health education messages (such as the dangers of smoking or the importance of health checks) which could have a beneficial influence on their health through health promoting behaviours and life styles (Fuchs, 1979; Keneckel, 1991). There is some evidence that greater education is associated with lower rates of smoking (Winkleby *et al*., 1992) and regular health checks (Coburn and Clyde, 2002).

(f) A person's *sense of control* over their life may mediate the association between education and health. Education increases the sense of personal control by developing analytic and communication skills (Mirowsky and Ross, 1998). The sense of personal control may also improve health through enhancing healthy behaviours, by controlling one's immediate addiction for future long-term health benefits (Folkman, 1984), whereas the lack of personal control may be a stressor with consequent adverse physiological consequences (Ross and Wu, 1995).

No study to date has considered all six of these pathways simultaneously. The six pathways are shown graphically in Fig. 1. The childhood variables, cognitive ability, father's social class and childhood and adolescent health, could affect both adult health and educational attainment, thus potentially confounding the association between education and health. If each pathway were analysed separately, the aim would be to establish whether the childhood variable explains the pathway that is marked with a question-mark, i.e. the association between education and adult health. Similarly, the last three pathways suggest that adult social class, sense of control and healthy behaviours could mediate the effect of education on health. Hence, each of the dotted pathways from education to adult health are also labelled with a question-mark, because any one of these variables may explain the association between education and health.

The chronological ordering of the variables making up the six pathways is shown in Fig. 2. Variables from early life (at age 7 years) are at the extreme left-hand end of Fig. 2 because they precede events and conditions in later life, such as health in adolescence (age 16 years), educational qualifications (age 23 years), adult social class, health behaviours and sense of control (all at age 33 years) and adult health (age 42 years).

Fig. 3 shows the pathway diagram that is formed by combining the pathways (bold lines) from Fig. 1. In addition to the arrows representing the six pathways, we also allow variables that are measured after childhood to depend on variables which precede them chronologically. These are denoted by arrows with dotted rather than bold lines. We also allow for associations between variables that are measured at the same time and indicate these by using dotted lines without arrow-heads. Dotted lines are used to emphasize associations between the variables which are not hypothesized to drive the education–health relationship, where failure to adjust for these associations would lead to biased estimates of each causal pathway.

Fig. 2. Time line of pathway mechanisms underlying the association of education and health over the life course

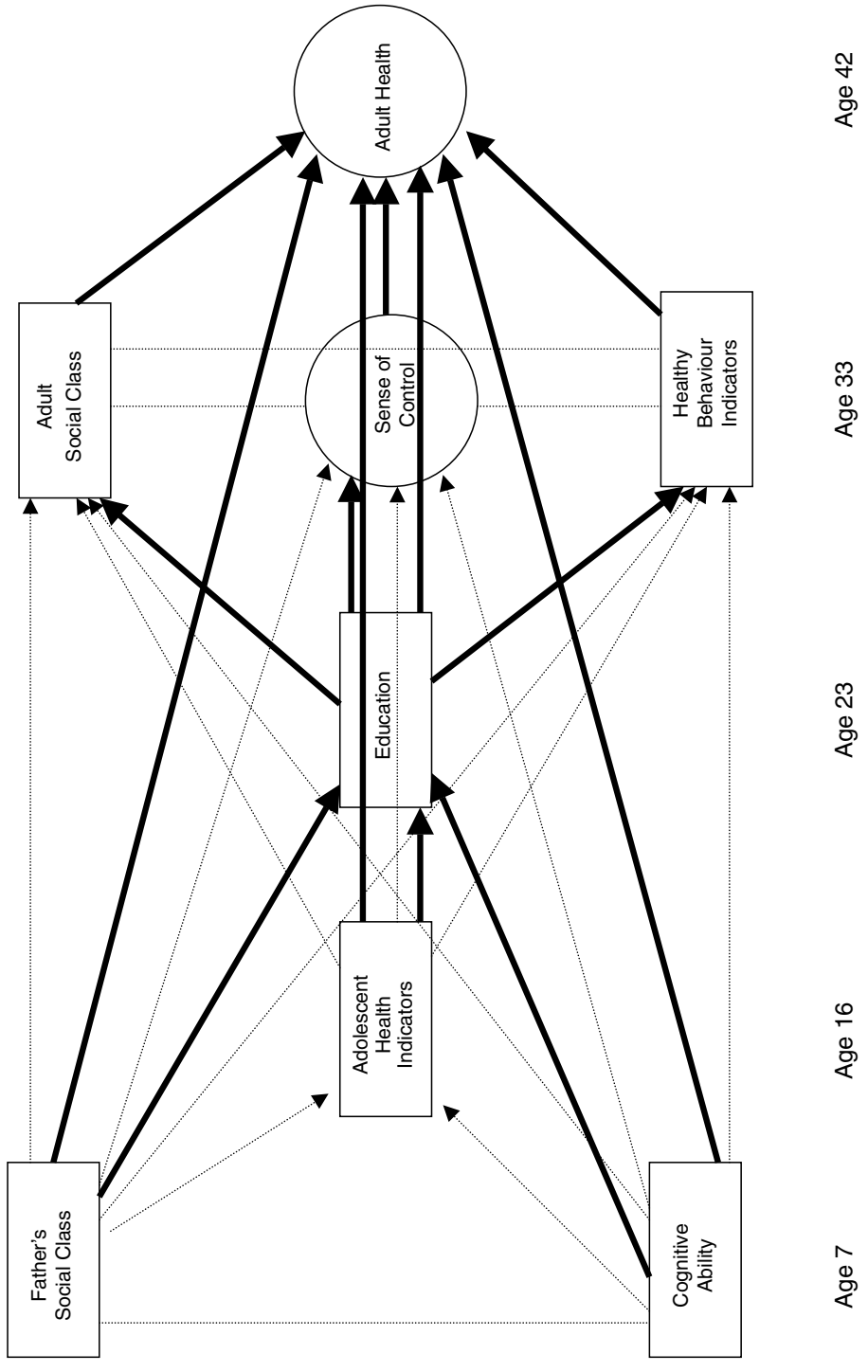

Fig. 3. Causal pathway model of the association between education and health

3. National Child Development Study

The National Child Development Study (NCDS) is a longitudinal study which takes as its subjects all those living in Great Britain who were born in the week March 3rd–9th, 1958 (Ferri, 1993). Following the initial 1958 survey, there have been six attempts to trace all members of the birth cohort and to monitor their physical, educational and social development. These were carried out in 1965 (at age 7 years), 1969 (age 11 years), 1974 (age 16 years), 1981 (age 23 years), 1991 (age 33 years) and 1999 (age 42 years). The NCDS is a unique, publicly available, source of data containing measures on individual characteristics and background information from different stages of the life course, which is essential if we are to model the hypothesized mechanisms. More precisely, we need measures of the following groups: adult and adolescent health status, sense of control, healthy behaviours, education, childhood and adult social class, and childhood cognitive ability.

The distributions of the variables are given in Table 1 and Fig. 4.

3.1. Adult health

Current adult health is taken to be that at age 42 years. We chose four variables from the NCDS related to adult health: self-rated health in general, self-rated health over the last 12 months, limiting (in terms of paid work) long-standing illness and hospital admission (excluding accidents).

Rather than consider each of the four adult health variables separately, we chose instead to follow Mirowsky and Ross (1998), Sandana *et al*. (2002) and Singh Manoux *et al*. (2002) and to analyse a single latent variable 'adult health'. The latent variable approach is convenient because it allows health status, which is inherently unobservable, to be represented by a number of indicators representing the underlying health state(s). In Section 4, we describe how to specify the adult health latent variable by using the four health variables in the measurement model component of a structural equation model (SEM).

3.2. Adolescent health

Adolescent health at age 16 years was represented by two indicators: frequency of school absences due to ill-health in the previous year and the presence of a handicap or disability. These were recoded so that a higher value indicated better health. A measure of adolescent health was chosen over child health (earlier in life) because of the former's stronger association with educational attainment.

3.3. Sense of control

The NCDS has three variables relating to 'sense of control', at age 33 years, which are combined into a single latent variable (see Section 4). Respondents were asked whether they had free choice and control over their lives, whether they got what they wanted out of life and whether they could run their lives as they wanted to (Ford *et al*., 2004). The (binary) questions were coded such that higher values indicated a higher sense of control.

3.4. Healthy behaviours

Three indicators of healthy behaviours at age 33 years were chosen—regular exercise, fruit consumption and smoking. These were coded such that a higher value indicated healthier behaviours.

Table 1. Distribution of all variables: 17416 cases in the analysis

Region (age 0 years)	
England	14517
Wales	914
Scotland	1985
Gender (age 0 years)	
Male	8999
Female	8412
Missing	5
Low birth weight (age 0 years)	
No	13931
Yes	3307
Missing	178
Father's social class (age 7 years)	
Professional	923
Managerial	2419
Skilled non-manual	1698
Skilled manual	8060
Partly skilled	2905
Unskilled	1326
Missing	85
Adolescent health age 16 years	
School absences last year—ill-health	
1, over 3 months	211
2, 1–3 months	831
3, 1 week–1 month	3784
4, less than 1 week	5894
Missing	6696
Is child handicapped or disabled?	
0, yes	786
1, no	9487
Missing	7143
Education and social class	
Highest qualifications gained (age 23 years)†	
0, no qualification	1436
1, 2–5 CSEs or equivalent NVQ1	1382
2, O-level or equivalent NVQ2	3660
3, A-level or equivalent NVQ3	1748
4, higher qualification NVQ4	1000
5, degree or higher NVQ5,6	911
Missing	7279
Social class (age 33 years)	
Professional	498
Managerial	2717
Skilled non-manual	2641
Skilled manual	1929
Partly skilled	1966
Unskilled	558
Missing	7107
Sense of control age 33 years	
Usually have control over life	
0, no	1195
1, yes	8863
Missing	7358
Never get what I want of life	
0, no	2482
1, yes	7593
Missing	7341
Satisfaction with way I run my life	
0, no	670
1, yes	9383
Missing	7363
Healthy behaviours age 33 years	
Smokes cigarettes	
0, yes	3582
1, no	7208
Missing	6626
Regular sport or exercise	
0, no	2363
1, yes	8406
Missing	6647
How often eats fresh fruit	
1, never	368
2, less than once a week	1091
3, 1 or 2 days a week	2060
4, 3–6 days a week	1671
5, once a day	3723
6, more than once a day	1891
Missing	6612

(*continued*)

Table 1 (*continued*)

Adult health age 42 years			
How is your health generally		Health over the last 12 months	
1, poor	382	1, poor	1061
2, fair	1566	2, fair	1715
3, good	5587	3, good	5007
4, excellent	3251	4, excellent	3004
Missing	6630	Missing	6629
Limiting long-standing illness		Hospital admissions—excluding accidents	
0, limiting health	1284	0, hospital admission	3637
1, no health problems	9494	1, no admissions	7145
Missing	6638	Missing	6634

†CSE, Certificate of Secondary Education; NVQ, National Vocational Qualification.

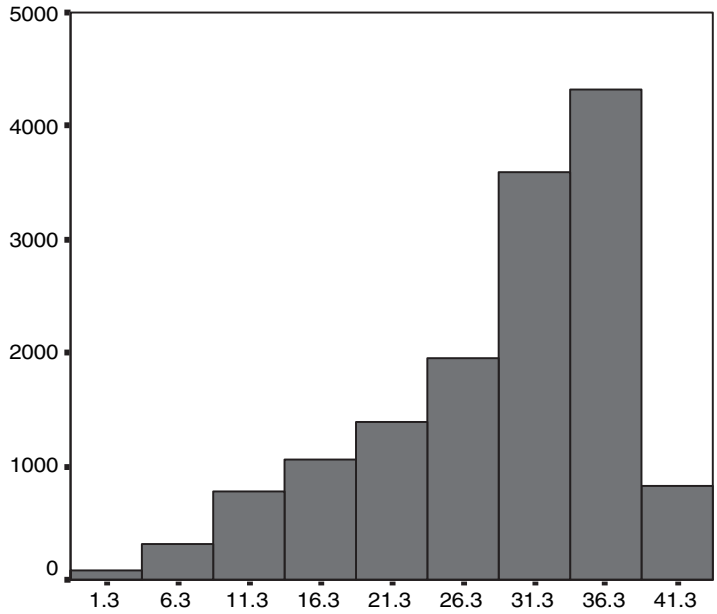

Fig. 4. Cognitive ability distribution (reading and mathematics scores), age 7 years (standard deviation 8.71; mean 28.5; $N = 14327.00$)

3.5. Education, social class and childhood cognitive ability

Education at age 23 years was measured by a six-category ordinal measure of increasing academic or vocational qualifications from no qualifications to degree level or higher. Father's social class (father's social class at age 7 years, supplemented by father's social class at birth if missing) and adult social class (at age 33 years) were measured by the Registrar General's social classification, an ordinal measure ranging from 1 (professional occupations) to 6 (manual unskilled occupations). Childhood cognitive ability (at age 7 years) was measured by summing the mathematics and reading test scores so that a higher score indicates greater ability. The distribution is shown in Fig. 4. Ability tests at age 7 years were chosen over intelligence tests (measured at age 11 years in the NCDS) to reduce the potential for the confounding of family and school effects on intelligence.

4. Data analysis using structural equation models

4.1. Why structural equation models?

Most previous studies of education and health were based on regression models. For example, Davey Smith *et al*. (1998) used proportional hazards regression to compare disease risk predictions by using education groups adjusted for adult occupational status. Berger and Leigh (1989) used instrumental variable regression models to estimate the direct effect of education on health while adjusting for

(a) endogeneity due to unobserved or unobservable factors influencing both education and health, and
(b) reverse causality whereby childhood health determines educational attainment and induces a spurious association between education and current health.

The study by Ross and Wu (1995) is similar to ours, in that a hypothesis was considered whereby education determines health through work and economic conditions in later life, which in turn give rise to better social–psychological resources, which in turn result in better health-related life style behaviours. Their analysis consisted of fitting four regression models in succession to establish whether this causal sequence explained the association between education and health.

Following Ross and Wu (1995) by fitting successive regression models is a valid way to proceed if the causal hypothesis that is being studied is relatively simple. When the causal hypothesis is more complex, such as that in Fig. 3, a more sophisticated approach is required. Hence, we used SEMs to analyse these data. The suitability of SEMs for studying education and health has already been demonstrated by Mirowsky and Ross (1998), who used them to test three pathways that were specific to the human capital hypothesis using data based on retrospective questions from a cross-sectional sample of the US population. SEMs can be used to specify 'causal models' through which causal hypotheses about the relationships between variables can be constructed and tested (Pearl, 1995; Greenland, 2000; Sobel, 2000; Greenland and Brumback, 2002). Moreover, 'indirect' effects are straightforward to estimate by using SEMs. For an example of an indirect effect, consider pathway (f) in Section 2; the indirect effect of education on health operates through a sense of control, because education affects an individual's sense of control in later life, which determines (in part) their adult health. There are two other indirect effects of education on adult health, operating through adult social class (pathway (d)) and healthy behaviours (pathway (e)). Pearl (2001) reiterated the point that effective policy interventions typically must account for indirect effects between variables, and not just direct effects. The total indirect effect of education on adult health is the sum of these three indirect pathways.

4.2. About structural equation models

SEMs are an extension of standard regression models to include multiple outcomes, called 'endogenous variables', and unobservable 'latent' variables (Muthén, 2004). The first component of an SEM is its structural model. For each endogenous variable there is a corresponding regression equation, which can depend on other endogenous variables as well as 'exogenous' variables; exogenous variables are equivalent to the predictor variables, or covariates, in standard regression modelling.

The second component of an SEM is its measurement model. It is appropriate to specify a measurement model when the observed study variables are believed to be imperfect indicator variables for an unobserved (or unobservable) underlying construct that is represented by one or more latent variables. Sense of control is an example of a variable which is best represented

by a latent variable. If the underlying construct is univariate, the simplest form of measurement model consists of indicator variables that are related to a single latent variable via a set of regression equations, one for each indicator, in which the indicator variable is the outcome and the latent variable is the predictor. Both the measurement and the structural models include normally distributed error terms which can be specified to be correlated or uncorrelated with each other. Together, the observed exogenous, endogenous and indicator variables are called 'manifest' variables.

We used the software package Mplus (Muthén and Muthén, 2003) to specify and fit SEMs for observed variables with interval, dichotomous and ordinal measurement scales. Broadly speaking, specifying an SEM involves setting

(a) the manifest variables' measurement scale,
(b) which variables are exogenous, endogenous, indicator and latent,
(c) the endogenous and exogenous variables in each structural model,
(d) the indicator variables in each measurement model and
(e) the error structure.

All error terms were set to be uncorrelated throughout this paper.

We now describe SEMs for the pathway diagrams in Figs 1 and 3. The variables that are denoted by circles are latent variables η_i (sense of control and adult health), with squares denoting either exogenous background variables x_i (cognitive ability score and father's social class) or endogenous outcome variables z_i (adolescent health, education, adult social class and healthy behaviours). Strictly speaking, latent variables can also be endogenous, in that they are outcomes in regression equations, but the distinction is made to identify latent variables and their associated measurement model. To aid clarity, each circle represents a latent variable and its measurement model. For example, adult health is a latent variable with a measurement model based on four indicators measuring self-rated health (two items), limiting long-standing illness and hospital admissions. Furthermore for clarity, the square for adolescent health represents two observed variables (school absences and disability), and the healthy behaviours variable represents three variables (exercise, fruit consumption and smoking).

4.2.1. Measurement models

The measurement model for adult health is given by four equations:

$$
\begin{aligned}
z_i^{\mathrm{GSRH}} &= \nu_1 + \lambda_1 \eta_i^{\mathrm{AH}} + \varepsilon_{1i},\\
z_i^{\mathrm{SRH12}} &= \nu_2 + \lambda_2 \eta_i^{\mathrm{AH}} + \varepsilon_{2i},\\
z_i^{\mathrm{LLSI}} &= \nu_3 + \lambda_3 \eta_i^{\mathrm{AH}} + \varepsilon_{3i},\\
z_i^{\mathrm{AHA}} &= \nu_4 + \lambda_4 \eta_i^{\mathrm{AH}} + \varepsilon_{4i},
\end{aligned}
$$

where the latent variable is $\eta_i^{\mathrm{AH}} =$ adult health, and the observed variables, which were not included in Figs 1–3, are $z_i^{\mathrm{GSRH}} =$ general self-rated health (ordinal scale), $z_i^{\mathrm{SRH12}} =$ self-rated health in the last 12 months (ordinal), $z_i^{\mathrm{LLSI}} =$ limiting long-standing illness (binary) and $z_i^{\mathrm{AHA}} =$ adult hospital admissions (binary). The error terms ε are independent and normally distributed.

As an important aside, note that, for notational simplicity, the model equations above and throughout this paper all involve a slight abuse of notation. The left-hand side of any model equation with a dichotomous or ordinal outcome represents a continuous latent variable rather than the manifest variable itself, where the latent and manifest variables are functionally linked. For example, take the binary limiting long-standing illness variable; the left-hand side of the

model equation represents a continuous latent variable that is positive for people with a limiting long-standing illness and negative otherwise. Moreover, the error term is constrained to have unit variance. Details of this functional link are suppressed in the above expression. See Múthen and Múthen (2003) for further details on the treatment of binary and ordinal variables.

Continuing with the description of the model parameters, the λ-parameters are referred to as 'factor loadings'. The interpretation of a factor loading depends on the measurement scale of the outcome variable. If the outcome variable is binary, λ is a probit regression coefficient; if the outcome variable is ordinal, λ is an ordinal probit regression coefficient. A similar set of equations determines the measurement model for the other variable, sense of control. These measurement models are used in every SEM which includes the adult health and sense of control variables.

4.2.2. Structural models

The structural models for the pathway diagrams in Figs 1 and 3 involve specifying a predictor equation for every variable with at least one single-headed arrow pointing to it. For example, the structural model corresponding to Fig. 1(a) is

$$\begin{aligned} z_i^{\mathrm{EDU}} &= \alpha_0 + \gamma_{01} x_i^{\mathrm{COG}} + \varepsilon_{i1}, \\ \eta_i^{\mathrm{AH}} &= \alpha_1 + \gamma_{11} x_i^{\mathrm{COG}} + \gamma_{12} z_i^{\mathrm{EDU}} + \varepsilon_{i2} \end{aligned}$$

where η_i^{AH} is adult health from the measurement model above, z_i^{EDU} is the ordinal education variable and x_i^{COG} is the cognitive test score. Noting again the comments above about binary and ordinal manifest variables, the α- and γ-parameters are referred to as regression coefficients. The interpretation of a regression coefficient depends on the measurement scale of the outcome variable. If the outcome variable is latent, α and γ are linear regression coefficients; if the outcome variable is binary, α and γ are probit regression coefficients; if the outcome variable is ordinal, α and γ are ordinal probit regression coefficients.

The two equations together are a simultaneous system, where γ_{01} is the (unadjusted) effect of cognition on education and γ_{12} is the effect of education on health adjusted for cognitive score. We make use of these interpretations in the 'adjusted' analyses of Sections 4.4 and 5. The structural models for Figs 1(b)–1(f) are similarly constructed.

The more complex structural model for Fig. 3 is (noting again the abuse of notation for binary and ordinal variables)

$$\begin{aligned} z_i^{\mathrm{TH}} &= \alpha_0 + \gamma_{01} x_i^{\mathrm{FSC}} + \gamma_{02} x_i^{\mathrm{COG}} + \varepsilon_{0i}, \\ z_i^{\mathrm{EDU}} &= \alpha_1 + \gamma_{11} x_i^{\mathrm{FSC}} + \gamma_{12} x_i^{\mathrm{COG}} + \gamma_{13} z_i^{\mathrm{TH}} + \varepsilon_{1i}, \\ z_i^{\mathrm{ASC}} &= \alpha_2 + \gamma_{21} x_i^{\mathrm{FSC}} + \gamma_{22} x_i^{\mathrm{COG}} + \gamma_{23} z_i^{\mathrm{TH}} + \gamma_{24} z_i^{\mathrm{EDU}} + \gamma_{25} z_i^{\mathrm{HB}} + \gamma_{26} \eta_i^{\mathrm{SOC}} + \varepsilon_{2i}, \\ z_i^{\mathrm{HB}} &= \alpha_3 + \gamma_{31} x_i^{\mathrm{FSC}} + \gamma_{32} x_i^{\mathrm{COG}} + \gamma_{33} z_i^{\mathrm{TH}} + \gamma_{34} z_i^{\mathrm{EDU}} + \gamma_{36} \eta_i^{\mathrm{SOC}} + \gamma_{37} z_i^{\mathrm{ASC}} + \varepsilon_{3i}, \\ \eta_i^{\mathrm{SOC}} &= \alpha_4 + \gamma_{41} x_i^{\mathrm{FSC}} + \gamma_{42} x_i^{\mathrm{COG}} + \gamma_{43} z_i^{\mathrm{TH}} + \gamma_{44} z_i^{\mathrm{EDU}} + \gamma_{45} z_i^{\mathrm{HB}} + \gamma_{47} z_i^{\mathrm{ASC}} + \varepsilon_{4i}, \\ \eta_i^{\mathrm{AH}} &= \alpha_5 + \gamma_{51} x_i^{\mathrm{FSC}} + \gamma_{52} x_i^{\mathrm{COG}} + \gamma_{53} z_i^{\mathrm{TH}} + \gamma_{54} z_i^{\mathrm{EDU}} + \gamma_{55} z_i^{\mathrm{HB}} + \gamma_{56} \eta_i^{\mathrm{SOC}} + \gamma_{57} z_i^{\mathrm{ASC}} + \varepsilon_{5i}, \end{aligned}$$

where z_i^{ASC} = adult social class (ordinal), η_i^{SOC} = sense of control (latent) and x_i^{FSC} = father's social class (exogenous). The equation for z_i^{TH}, adolescent health, actually represents two equations: one for handicap (binary) and another for school absences due to ill-health (ordinal). Similarly, the equation for z_i^{HB}, health behaviours, represents three equations: smoking (binary), exercise (binary) and fruit consumption (ordinal). None of the error terms are correlated.

4.3. Parameter estimation and missing data

Parameter estimation is performed by maximum likelihood estimation in Mplus. If data on the outcomes are missing, Mplus can perform 'full information' maximum likelihood estimation using the incomplete data (i.e. the complete and incomplete cases) provided that

(a) the data can be assumed to be either missing completely at random or missing at random (Little and Rubin, 2002) and
(b) the latent outcomes and latent variables are normally distributed.

Assumption (b) corresponds to the standard SEM assumption and, as such, is not restrictive. However, in most analyses the assumption of missingness completely at random is unlikely to hold; the missingness at random assumption is more realistic if factors that are strongly associated with non-response or drop out are included in the model. In the NCDS data, out of the 17416 babies in the initial birth cohort, 10979 adults remained in the cohort at age 42 years, which corresponds to an attrition rate of around 37%. Hawkes and Plewis (2006) analysed patterns of attrition among the NCDS cohort. The results suggest that being born in Wales, having a mother with lower education and low birth weight significantly predict drop-out. Including these variables in the model as covariates makes the assumption of missingness at random more plausible. As such, they are included in each of the following SEMs as background exogenous variables (although they do not appear in the diagrams). If any exogenous variable is missing, similar arguments to those above hold provided that the additional assumption is made that

(c) the distribution of the missing exogenous variable, given the other variables, is normal.

4.4. Analytical strategy

The analysis that is presented in Section 5.2 is broken into three parts.

(a) *Unadjusted analysis*: estimate the associations between adult health and education, and between adult health and each of the pathway variables that were identified in Section 2. This is carried out by using a series of SEMs, one for each pair of variables. The first SEM corresponds to a pathway diagram with a single-headed arrow drawn from education to adult health, the second model corresponds to a pathway diagram with a single-headed arrow from cognitive ability to adult health, and so on. This is used to establish unadjusted associations between the variables and adult health.
(b) *Adjusted analysis*: estimate the association between education and adult health, adjusted for each of the variables in turn. This is performed by using another series of SEMs for the pathways in Fig. 1, specified in the manner that was described above. This is used to establish whether the association between education and adult health remains after adjustment.
(c) *Causal analysis*: the full set of direct and indirect effects of education on health were estimated by using the full SEM that was specified for Fig. 3 in Section 4.3.

Model fit is assessed by using a series of fit indices that measure how closely the estimated covariance matrix for an SEM fits the sample associations between the observed endogenous, exogenous and indicator variables. It is standard to use several criteria to assess fit because no index has been shown to perform well in all situations. For instance, the χ^2-statistic measures the absolute goodness of fit but can be sensitive to minor model misspecifications when sample sizes are large. The root-mean-square error of approximation RMSEA gives a measure of the discrepancy in fit per degrees of freedom; it is 0 if the model fits exactly. The comparative fit index CFI has values between 0 and 1 with higher values indicating a better fit. Hu and Bentler

(1999) recommended the following fit index cut-off to judge a 'good'fit: Tucker Lewis index TLI > 0.95, comparative fit index CFI > 0.95, root-mean-squared error of approximation RMSEA < 0.06 and weighted root-mean-square residual WRMR < 0.90.

The analyses were performed separately for men and women by using MPlus. Both complete and incomplete cases were included in the analysis, using full information maximum likelihood estimation to fit the model.

5. Results

5.1. Measurement model

Two measurement models for the latent variables adult health and sense of control in the analysis were specified as described in Section 4.3. Each of the factor loadings are displayed in Table 2. There was some evidence for unidimensionality of these constructs. Cronbach's α-coefficient for adult health was 0.67 and for sense of control was 0.63. Furthermore, the percentage of the total variance explained by the first principal component for both the adult health and sense of control variables is over 60%. The goodness-of-fit statistics for adult health indicate that the measurement model fits well. Goodness-of-fit statistics are not applicable for sense of control because there were only three indicators. However, each of the factor loadings is significant, suggesting that there is common variance between all three indicators for both latent variables.

5.2. Structural equation modelling

5.2.1. Men

5.2.1.1. Unadjusted analyses. The associations between adult health and education, and between adult health and the other variables in Fig. 1 are shown in the second column of Table 3.

Table 2. Factor loadings and goodness-of-fit statistics of the latent variables adult health and sense of control

	Results for women	*Results for men*
(Good) health at age 42 years by		
Good health in general	0.95†	0.93†
Good health over the past year	0.92†	0.95†
No limiting long-standing illness	0.69†	0.62†
No hospital admissions (not accidents)	0.20†	0.18†
χ^2, degrees of freedom	3.4, 1	10.7, 1
Comparative fit index CFI	1.00	1.00
Tucker Lewis index TLI	0.99	1.00
Root-mean-square-error of approximation RMSEA	0.02	0.04
Weighted root-mean-square residual WRMR	0.48	0.86
(Lack of) sense of control at age 33 years by		
Usually have no choice and control over life	0.85†	0.80†
Never get what I want out of life	0.88†	0.84†
Find life's problems too much for me	0.85†	0.81†
χ^2, degrees of freedom	0.00, 0	0.00, 0
Comparative fit index CFI	1.00	1.00
Tucker Lewis index TLI	1.00	1.00
Root-mean-square error of approximation RMSEA	NA‡	NA‡
Weighted root-mean-square residual WRMR	0.00	0.00

†$p < 0.01$.
‡NA, not applicable.

Table 3. Determinants of adult health: univariate, adjusted and causal analysis—standardized regression coefficients†

Variable	*Coefficients from the following analyses:*											
	Unadjusted analysis	*Adjusted analysis*									*Causal analysis*	
											Direct	*Indirect*
Men												
Highest qualifications age 23 years	0.20‡	0.18‡	0.20‡	0.17‡	0.18‡	0.14‡	0.09‡	0.12‡	0.21‡	0.19‡	−0.04	0.17‡
Cognitive ability age 7 years	0.16‡	0.07‡									0.02	0.13‡
Father's social class age 7 years	−0.11‡		−0.03§								0.00	−0.09‡
No school absences age 16 years	0.15‡			0.12‡							0.08‡	0.05‡
No handicap or disability age 16 years	0.13‡				0.11‡						0.16‡	0.03§
Social class age 33 years	−0.19‡					−0.13‡					−0.07§	NA§§
Sense of control age 33 years	0.35‡						0.32‡				0.23‡	NA§§
Non-smoker age 33 years	0.22‡							0.26‡			0.21‡	NA§§
Regular exercise age 33 years	0.02								0.05		0.06§	NA§§
Fresh fruit age 33 years	0.10‡									0.07‡	−0.01	NA§§
Women												
Highest qualifications age 23 years	0.23‡	0.22‡	0.23‡	0.22‡	0.24‡	0.21‡	0.16‡	0.17‡	0.24‡	0.20‡	0.03	0.11‡
Cognitive ability age 7 years	0.18‡	0.07‡									0.03	0.12‡
Father's social class age 7 years	−0.15‡		−0.06‡								−0.03	−0.11‡
No school absences age 16 years	0.18‡			0.12§							0.11‡	0.04‡
No handicap or disability age 16 years	0.11‡				0.09‡						0.16‡	0.02
Social class age 33 years	−0.14‡					−0.13§					0.01	NA§§
Sense of control age 33 years	0.31‡						0.26‡				0.20‡	NA§§
Non-smoker age 33 years	0.20‡							0.20‡			0.14‡	NA§§
Regular exercise age 33 years	0.11‡								0.16‡		0.03	NA§§
Fresh fruit age 33 years	0.16‡									0.12‡	0.05§	NA§§

†Unadjusted analysis: all models are of the form adult health = variable. Adjusted analysis: all models are of the form adult health = variable + X, where X denotes cognitive ability, father's social class, adolescent health indicators, sense of control, adult social class and healthy behaviour indicators.
‡$p < 0.01$.
§$p < 0.05$.
§§NA, not applicable.

The second column displays the coefficients of the probit regression of cognitive score on adult health, the regression of father's social class on adult health etc. To compare the magnitude of the effects, we present standardized regression coefficients which take values from −1 to 1, with positive values indicating positive association and 0 indicating no association.

All the variables were significantly associated with adult health, and all these associations were in the direction expected. Boys with higher scores on the arithmetic and reading tests (i.e. who had higher cognitive ability) had better adult health. Those whose fathers had worked in manual social class occupations had poorer adult health. Those with better adolescent health had better adult health. Men with higher educational qualifications at age 23 years had better health at age 42 years. Men with a higher sense of control at age 33 years had better health. Those who worked in manual social class occupations at age 33 years had poorer adult health. Men with healthier behaviours at age 33 years had better adult health.

5.2.1.2. Adjusted analyses. The next nine columns of Table 3 contain the associations between education and adult health, adjusted for each of the other variables separately. The unadjusted association (0.20) reduces slightly after adjusting for cognitive ability (0.18), school absences (0.17), handicap (0.18), adult social class (0.14), sense of control (0.09), (non-)smoking (0.12) and fresh fruit consumption (0.19) but not for father's social class and regular exercise. The largest reduction in the effect of education on adult health occurs after adjustment for sense of control and for (non-)smoking.

5.2.1.3. Causal analysis. When fitting the full SEM for Fig. 3 (Table 3), the direct effect of education on adult health was not significantly different from 0. However, the indirect effect of education on adult health (through its effect on adult social class, control and health behaviours) is positive and significant. Most of this indirect effect is through sense of control and (non-)smoking (Table 4). All the early life variables (ages 7–16 years) indirectly affect adult health (Table 3), with educational qualifications being the most important mediator (Table 4). None of the variables at age 33 years were specified to have indirect effects on adult health (see Fig. 3).

The goodness-of-fit statistics for the full SEM in the causal analysis columns (Tables 3 and 4) indicate a good fit. Although the χ^2-statistic is large (208.026 on 56 degrees of freedom), the CFI- (1.00), TLI- (0.99), RMSEA- (0.02) and WRMR- (0.90) statistics are within acceptable limits.

5.2.2. Women

A similar picture was obtained when analysing the pathways between education and health for women (Table 4). In the unadjusted analyses, the association of education with adult health is positive and statistically significant. The associations with adult health of all the other variables are also statistically significant and in the expected directions. In the adjusted analyses, the effect of education on adult health is somewhat reduced when adjusted for sense of control, (non-)smoking and fresh fruit consumption.

In the causal analyses, the direct effect of education on adult health is not significantly different from 0, whereas the indirect effect is significant and positive. However, the indirect effect of education on adult health (through its effect on adult social class, control and health behaviours) is positive and significant. Like in men, most of this indirect effect is through sense of control and (non-)smoking (Table 4). All the early life variables (ages 7–16 years) indirectly affect adult health (Table 3), with educational qualifications being the most important mediator (Table 4).

Table 4. Indirect effects of education, father's social class, cognitive ability and adolescent health indicators on adult health: standardized regression coefficients

Variable	*Regression coefficients for men*	*Regression coefficients for women*
Indirect effects of education (age 23 years) on health (age 42 years)		
Pathways via		
Social class age 33 years	0.03†	−0.01
Sense of control age 33 years	0.07†	0.05†
Smoking age 33 years	0.07†	0.05†
Exercise age 33 years	0.01‡	0.01‡
Fruit consumption age 33 years	0.00	0.01
Indirect effects of school absences (age 16 years) on health (age 42 years)		
Pathways via		
Qualifications age 23 years	0.02†	0.03†
Social class age 33 years	0.01	0.00
Sense of control age 33 years	0.02	0.01‡
Smoking age 33 years	0.02†	0.02†
Exercise age 33 years	0.01‡	0.00
Fruit consumption age 33 years	0.00	0.00
Indirect effects of being handicapped (age 16 years) on health (age 42 years)		
Pathways via		
Qualifications age 23 years	0.01	0.01
Social class age 33 years	0.00	0.00
Sense of control age 33 years	0.04†	0.02
Smoking age 33 years	−0.01	0.00
Exercise age 33 years	0.00	0.00
Fruit consumption age 33 years	0.00	0.00
Indirect effects of father's social class (age 7 years) on health (age 42 years)		
Pathways via		
School absences age 16 years	−0.02†	−0.02†
Being handicapped aged 16 years	−0.01	−0.02†
Qualifications age 23 years	−0.04†	−0.05†
Social class age 33 years	−0.02†	0.00
Sense of control age 33 years	−0.03†	−0.03†
Smoking age 33 years	−0.03†	−0.03†
Exercise age 33 years	0.00	0.00
Fruit consumption age 33 years	0.00	−0.01‡
Indirect effects of cognitive ability (age 7 years) on health (age 42 years)		
Pathways via		
School absences age 16 years	0.02†	0.02†
Being handicapped aged 16 years	0.05†	0.05†
Qualifications age 23 years	0.06†	0.06†
Social class age 33 years	0.02†	0.00
Sense of control age 33 years	0.05†	0.04†
Smoking age 33 years	0.02†	0.02†
Exercise age 33 years	0.01‡	0.00
Fruit consumption age 33 years	0.00	0.01‡

†$p < 0.01$.
‡$p < 0.05$.

Table 5. Correlations between all the variables in the causal analysis†

	Health in general	*Health last year*	*Limiting health*	*Hospital admissions*	*School absence*	*Handi-cap*	*Control over life*	*Get what I want*	*Satisfied with life*	*Eat fruit*	*Not smoking*	*Exer-cise*	*Father's class*	*Qualifi-cations*	*Social class*	*Mother's education*	*Low birth weight*	*Wales*	*Scot-land*	*Cognitive score*
Health in general		0.88	0.66	0.17	0.18	0.24	0.22	0.27	0.28	0.17	0.26	0.13	−0.16	0.26	−0.14	0.11	−0.05	0.01	−0.05	0.18
Health last year	0.88		0.63	0.20	0.17	0.19	0.21	0.24	0.27	0.14	0.24	0.11	−0.14	0.21	−0.12	0.11	−0.05	0.01	−0.03	0.15
No limiting health	0.58	0.58		0.16	0.22	0.25	0.13	0.17	0.25	0.14	0.17	0.09	−0.12	0.22	−0.13	0.09	−0.07	−0.03	−0.11	0.17
No hospital admission	0.14	0.17	0.19		0.08	−0.04	0.02	0.01	0.02	−0.01	0.09	−0.03	−0.01	0.00	0.00	0.00	0.00	0.02	−0.05	0.03
No school absence	0.17	0.16	0.23	0.01		0.17	0.11	0.08	0.11	0.06	0.18	0.06	−0.15	0.32	−0.13	0.11	−0.02	−0.07	−0.01	0.15
No handicap	0.22	0.21	0.26	−0.05	0.21		0.09	0.15	0.10	0.07	0.09	0.01	−0.12	0.22	−0.14	0.04	0.01	−0.02	0.06	0.26
Control over life	0.24	0.21	0.20	−0.04	0.12	0.17		0.75	0.72	0.14	0.17	0.15	−0.14	0.29	−0.20	0.12	0.00	0.01	−0.03	0.16
Get what I want	0.29	0.27	0.23	−0.02	0.10	0.17	0.67		0.75	0.17	0.28	0.16	−0.15	0.28	−0.20	0.14	−0.08	0.03	−0.03	0.18
Satisfied with life	0.29	0.26	0.30	0.01	0.17	0.25	0.65	0.68		0.16	0.17	0.19	−0.08	0.20	−0.17	0.09	−0.03	0.04	−0.04	0.17
Eat fruit	0.11	0.09	0.05	−0.03	0.06	0.08	0.11	0.13	0.09		0.26	0.24	−0.15	0.27	−0.15	0.11	−0.13	0.06	−0.06	0.14
Not smoking	0.30	0.27	0.18	0.01	0.14	0.02	0.13	0.24	0.21	0.24		0.18	−0.19	0.35	−0.21	0.08	−0.04	0.01	−0.13	0.15
Exercise	0.18	0.15	0.13	−0.06	0.14	0.08	0.15	0.14	0.17	0.25	0.23		−0.12	0.18	−0.10	0.12	−0.05	−0.02	−0.06	0.06
Father's class	−0.11	−0.09	−0.08	−0.01	−0.19	−0.06	−0.12	−0.15	−0.09	−0.06	−0.13	−0.06		−0.42	0.25	−0.36	0.08	0.06	0.10	−0.29
Qualifications	0.20	0.18	0.24	−0.02	0.30	0.20	0.30	0.30	0.25	0.14	0.32	0.22	−0.37		−0.48	0.36	−0.16	−0.02	0.04	0.52
Social class	−0.19	−0.17	−0.19	0.04	−0.18	−0.09	−0.21	−0.24	−0.13	−0.09	−0.24	−0.10	0.30	−0.49		−0.19	0.11	0.02	−0.03	−0.29
Mother's education	0.10	0.07	0.07	−0.03	0.14	0.04	0.11	0.08	0.09	0.07	0.03	0.07	−0.37	0.29	−0.22		−0.04	0.02	−0.07	0.20
Low birth weight	−0.08	−0.04	−0.07	−0.02	−0.04	−0.15	−0.02	0.00	−0.05	−0.07	−0.03	−0.07	0.08	−0.10	0.09	−0.09		0.00	−0.01	−0.17
Wales	0.01	0.02	−0.04	−0.01	−0.07	−0.06	0.05	0.00	0.07	0.00	0.01	−0.03	0.02	−0.06	0.05	0.07	−0.07		−0.58	−0.02
Scotland	−0.01	0.02	−0.05	0.02	−0.03	0.04	0.06	0.02	0.16	−0.03	−0.03	0.01	0.11	0.02	0.04	−0.12	0.01	−0.58		0.08
Cognitive score	0.16	0.14	0.17	−0.04	0.18	0.26	0.20	0.21	0.17	0.06	0.12	0.13	−0.28	0.51	−0.33	0.23	−0.14	−0.01	0.12	

†Correlations below the diagonal are for men; correlations above the diagonal are for women.

Although the χ^2-statistic for the full SEM in the causal analysis columns (Tables 3 and 4) is large (128.847 on 54 degrees of freedom), the other statistics indicate a good fit: CFI = 1.00, TLI = 1.00, RMSEA = 0.01 and WRMR = 0.70.

5.3. Summary of findings

In summary, the association between education and adult health appears to be explained by a combination of mechanisms. Most of the association between education and adult health is mediated through smoking and sense of control—educated men and women are more likely to be non-smokers and to have a higher sense of control, which in turn, is associated with better adult health. The correlations between the variables are shown in Table 5 both for men and for women.

6. Discussion

The results of this study are novel and represent a significant advance in the understanding of the association between education and health. In previous studies, it was found that education had a strong and positive direct effect on adult self-assessed health (Ross and Wu, 1995; White *et al*., 1999) and cardiovascular disease (Winkleby, 1992; Davey Smith *et al*., 1998) even after adjusting for potential mediators and confounders. In our analysis, the 'direct' association between educational qualifications and adult health did not behave in this way. In men and women, the direct effect of education on adult health was not significant. As such, we found little evidence for the hypothesis that increases in educational qualifications 'directly' translate into better health. This is similar to the findings by Mirowsky and Ross (1998), the only other study (to the authors' best knowledge) to analyse multiple pathways between education and health. This suggests that, when the multiple pathways between education and health are analysed by using appropriate methods, there is little evidence of a 'direct' or residual effect of education on health.

It is important to emphasize that the absence of evidence for a direct effect of education does not imply that education is unimportant for health. The indirect effect of education (through healthy behaviours and sense of control) was found to be positive. The study by Mirowsky and Ross (1998) also found that healthy life styles and a greater sense of control are important pathways between education and health.

Childhood cognitive ability (pathway (a) in Section 2) did not have any direct effect on adult health, although it had indirect effects and thus does not appear to confound the association between education and health. This echos the results of the Scottish Mental Health Survey of 1932 which found that adjusting for adult occupational social class attenuated the effects of intelligence on morbidity and mortality (Hart *et al*., 2003).

A greater perception of control (pathway (f)) at age 33 years was related to higher educational attainment, and better adult health, and thus mediates some of the association between education and health. There is evidence that perceptions of control in adulthood originate from higher parental education, higher cognitive ability and higher educational attainment (Lewis *et al*., 1999). It is possible that children with a greater sense of control go on to do better at school and have better educational, occupational and health outcomes. However, the NCDS data did not have any measures of sense of control in childhood or adolescence.

Good health in adolescence (pathway (c)) predicted higher educational attainment and better adult health, thus confounding some of the association between education and health. Although, in general, adolescents are biologically robust, with low morbidity and mortality, poor self-reported health at age 15 years is strongly associated with medically unexplained

physical symptoms at age 36 years (Hotopf *et al.*, 1999). Children with headaches are at an increased risk of recurring headache, multiple physical symptoms and psychiatric morbidity in adulthood (Fearon and Hotopf, 2001). The strong associations between adolescent health and adult health in this study are in line with these other findings.

Social class in adulthood (pathway (d)) had a direct effect on the health of adult men but not women. This gender difference is not unexpected—several other studies have found that, when women are assigned to a class on the basis of their own occupation, the association between social class and health is not strong (Chandola, 1998). In contrast, classifying women by their head of household's class results in stronger associations between social class and health. The indirect effect of women's father's social class (pathway (b)) on their adult health in this study also suggests that household social class is important for the health of women.

Although the health behaviours (pathway (e)) that were examined in this study were limited (self-reported exercise, fruit consumption and smoking) and crudely measured, healthy behaviours were strongly associated with better adult health. Smoking, in particular, was the strongest pathway linking education with health, mediating some of the association between the two. This suggests that to bring about improvements in health behaviours in the population we need to understand how people's behaviours are influenced by social and psychosocial factors throughout their life.

The research literature on education and health that has been reviewed in this paper has largely come from the disciplines of sociology and epidemiology. In economics, the human capital theory of health that was developed by Grossman (1972) also suggests a causal relationship from education to health, while conceding that the direction of causality may be reversed or that a 'third variable' that determines both education and health drives their correlation. The 'reverse causality' hypothesis is incorporated in this study by including a latent variable measure of adolescent health in the analysis. The third-variable hypothesis was also examined by including other possible confounders of the education–health association measures of childhood class and cognitive ability. However, there may be other variables that were not included in this analysis that could drive the association between education and health. Such variables from economics literature on the association between education and health include 'future orientation' (Fuchs, 1982) (i.e. people who have a high degree of time preference for the future attend school for longer periods and make greater investments in their health) and 'psychological capital' (Feinstein and Symons, 1999; Murasko, 2003), including locus of control, self-esteem, hostility and anxiety. Some of this psychological capital may be captured in our analysis through the variable 'sense of control'. It is possible that other variables, which were not included in our analysis, cause the relationship between education and health, although the set of (theroretically derived) variables that were examined in our analysis do a good job already in explaining the direct effects between education and health.

Another potential limitation of this study was its focus on British data. Country-specific cohort effects may have been observed in this study. For example, the establishment of comprehensive schools and new universities in the 1960s and 1970s may have uniquely affected the educational attainment of this cohort. Much of the literature that is referred to in this paper analyses British or US data. The association between education and health may differ in other countries, although the mechanisms underlying the association should be similar.

The statistical methods that were used in this paper were key to the extra insights that we have gained from this analysis. We made two important working assumptions when interpreting the SEM results. First is that the regression error terms in the model are normally distributed. Enders (2001) found SEM parameter estimates by using full information maximum likelihood to be robust to modest departures from normality (although standard error estimates could be

biased), but it is not possible to establish the true extent of non-normality when using latent variables or when the data are incomplete. Second is that no interactions were included in our analysis. Although it is possible that, say, the effect of sense of control on adult health may vary for different values of childhood cognition, no interactions were included because our focus was on existing hypotheses relating to education and adult health, none of which dealt with the role of moderating variables. However, it is easy to specify an SEM that includes interactions if such theories emerge.

With regard to the quality of data, although self-reported health measures are subject to various biases, the use of measurement (latent variable) models for adult health and sense of coherence enable better measures of these concepts than single-item measures. Mirowsky and Ross (1998) also used SEMs to address the education–health question, using retrospective questions from cross-sectional data from the USA. As such, our use of prospective data from the NCDS offers an improvement on their study that is commensurate with the benefits of prospective over retrospective measures. To offset this, there is the problem of missing data, from sample attrition, other drop-out and missing answers in completed questionnaires. However, the size of the sample responding in adulthood is in line with other high quality studies of adult health (Wadsworth *et al.*, 2003), and we have used estimation techniques which protect against non-response bias under a data 'missing at random' assumption (Little and Rubin, 2002). The possibility of non-response bias remains, however. For example, suppose that people with low intelligence and poor adolescent health were systematically missing from the study; although birth weight, education and other variables that are observed for (almost) everyone in the study would account for some non-response bias, there may still be systematic differences between the intelligence and adolescent health for those with the same birth weight, education, etc. who responded and for those who did not. We are thus forced to make the working assumption that any such biases are small.

Finally, note that we make no claims that our causal model represents the fundamental, true relationship between education and health. Rather, the term 'causal analysis' refers to the process by which widely believed theories are specified and assessed empirically. The analysis of these causal pathways in this study has shown that adolescent health, healthy behaviours and the sense of control contribute substantially to explaining this association. Furthermore, adult social class among men and father's social class among women also contribute to explaining the association between education and health. Policies that are aimed at improving adult population health and reducing social inequalities in health need to focus on the causal pathways that are highlighted in this study.

Acknowledgements

The data and tabulations that were used in this publication were made available through the Economic and Social Research Council Data Archive. The data were originally collected by the National Birthday Trust Fund, N. Butler (Perinatal Mortality Study), the National Children's Bureau (sweeps 1–4), City University, the Social Statistics Research Unit (sweep 5) and the Joint Centre for Longitudinal Research which comprises the Centre for Longitudinal Studies, the Institute of Education, University of London, the International Centre for Health and Society, University College Medical School, London, and the National Centre for Social Research—sweep 6. The National Centre for Social Research was responsible for assisting in the development of instrumentation, development of computer-assisted personal interviewing programs, conduct of fieldwork, initial data coding and editing, and some documentation. The Office for National Statistics co-ordinated funding on behalf of various UK Government depart-

ments. Neither the original collectors of the data nor the Data Archive bear any responsibility for the analyses or interpretations that are presented here.

The authors thank Mel Bartley, Fiona Steele and Archana Singh-Manoux for their helpful comments on earlier drafts. The authors are also grateful to the referees for their helpful comments. TC and PC were funded by an Economic and Social Research Council grant (RES-000-22-0290).

References

Batty, D. and Deary, I. (2004) Early life intelligence and adult health. *Br. Med. J.*, **329**, 585–586.
Berger, M. C. and Leigh, J. P. (1989) Schooling, self-selection, and health. *J. Hum. Resour.*, **24**, 433–455.
Blane, D. (2003) Explanations of the difference in mortality risk between different educational groups. *Int. J. Epidem.*, **32**, 355–356.
Blane, D., White, I. and Morris, J. (1996) Education, social circumstances and mortality. In *Health and Social Organization* (eds D. Blane, E. Brunner and R. Wilkinson). London: Routledge.
Chandola, T. (1998) Social inequality in coronary heart disease: a comparison of occupational classifications. *Socl Sci. Med.*, **47**, 525–533.
Coburn, D. and Clyde, R. (2002) Socioeconomic status and preventative health behavior. *J. Hlth Socl Behav.*, **15**, 67–78.
Dahl, E. (1994) Social inequalities in ill health: the significance of occupational status, education and income—results from a Norwegian study. *Sociol. Hlth Ill.*, **16**, 644–667.
Davey Smith, G., Hart, C. L., Hole, D. J., MacKinnon, P., Gillis, C., Watt, G., Blane, D. and Hawthorne, V. (1998) Education and occupational social class: which is the more important indicator of mortality risk? *J. Epidem. Commty Hlth*, **52**, 153–160.
Enders, C. K. (2001) The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data. *Psychol. Meth.*, **6**, 352–370.
Fearon, P. and Hotopf, M. (2001) Relation between headache in childhood and physical and psychiatric symptoms in adulthood: national birth cohort study. *Br. Med. J.*, **322**, 1145.
Feinstein, L. (2002) Quantitative estimates of the social benefits of learning, 2: health (depression and obesity). *Wider Benefits of Learning Research Report No. 6*. London: Centre for Research on the Wider Benefits of Learning.
Feinstein, L. and Symons, J. (1999) Attainment in secondary schools. *Oxf. Econ. Pap.*, **51**, 300–321.
Ferri, E. (1993) Life at 33: the fifth follow-up of the National Child Development Study. *Report*. National Children's Bureau, City University and Economic and Social Research Council, London.
Folkman, S. (1984) Personal control and stress and coping processes: a theoretical analysis. *J. Pers. Socl Psychol.*, **86**, 839–852.
Ford, R., Marsh, A. and Finlayson, L. (2004) *What Happens to Lone Parents*. London: Stationery Office.
Fuchs, V. (1979) Economics, health and post industrial society. *Millbnk Q.*, **57**, 153–182.
Fuchs, V. R. (1982) Time preference and health: an exploratory study. In *Economic Aspects of Health* (ed. V. R. Fuchs). Chicago: University of Chicago Press.
Gottfredson, L. S. (2004) Intelligence: is it the epidemiologists' elusive "fundamental cause" of social class inequalities in health? *J. Pers. Socl Psychol.*, **86**, 174–199.
Gottfredson, L. and Deary, I. (2004) Intelligence predicts health and longevity, but why? *Curr. Dir. Psychol. Sci.*, **13**, 1–4.
Greenland, S. (2000) Causal analysis in the health sciences. *J. Am. Statist. Ass.*, **95**, 286–289.
Greenland, S. and Brumback, B. (2002) An overview of relations among causal modelling methods. *Int. J. Epidem.*, **31**, 1030–1037.
Grossman, M. (1972) On the concept of health capital and the demand for health. *J. Polit. Econ.*, **80**, 223–255.
Hart, C. L., Taylor, M. D., Davey Smith, G., Whalley, L. J., Starr, J. M., Hole, D. J., Wilson, V. and Deary, I. J. (2003) Childhood IQ, social class, deprivation, and their relationships with mortality and morbidity risk in later life: prospective observational study linking the Scottish Mental Survey 1932 and the Midspan studies. *Psychsom. Med.*, **65**, 877–883.
Hawkes, D. and Plewis, I. (2006) Modelling non-response in the National Child Development Study. *J. R. Statist. Soc.* A, **169**, no. 3, in the press.
Hotopf, M., Mayou, R., Wadsworth, M. and Wessely, S. (1999) Childhood risk factors for adults with medically unexplained symptoms: results from a national birth cohort study. *Am. J. Psychiatr.*, **156**, 1796–1800.
Hu, L. and Bentler, P. M. (1999) Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equn Modlng*, **6**, 1–55.
Isohanni, I., Jobes, P. B., Jarvelin, M. R., Nieminen, P., Rantakallio, P., Jokelainen, J., Croudace, T. J. and Isohanni, M. (2001) Educational consequences of mental disorders treated in hospital: a 31-year follow-up of the Northern Finland 1966 Birth Cohort. *Psychol. Med.*, **31**, 339–349.

Jefferis, B. J., Power, C. and Hertzman, C. (2002) Birth weight, childhood socioeconomic environment, and cognitive development in the 1958 British birth cohort study. *Br. Med. J.*, **325**, 305.
Keneckel, D. (1991) Health behaviour, health knowledge and schooling. *J. Polit. Econ.*, **99**, 287–305.
Lewis, S., Ross, C. E. and Mirowsky, J. (1999) Establishing a sense of personal control in the transition to adulthood. *Socl Forces*, **77**, 1573–1599.
Little, R. J. A. and Rubin, D. B. (2002) *Statistical Analysis with Missing Data*, 2nd edn. Hoboken: Wiley.
Mirowsky, J. and Ross, C. E. (1998) Education, personal control, lifestyle and health: a human capital hypothesis. *Res. Agng*, **20**, 415–449.
Mirowsky, J. and Ross, C. E. (2003) *Social Status and Health*. New York: Aldine de Gruyter.
Morris, J. N., Blane, D. B. and White, I. R. (1996) Levels of mortality, education, and social conditions in the 107 local authority areas of England. *J. Epidem. Commty Hlth*, **50**, 15–17.
Murasko, J. (2003) The human capital model of health: childhood origins. *Working Paper*. Rice University, Houston. (Available from `http://www.ruf.rice.edu/~muraskoj/html/murasko_paper1.pdf`.)
Muthén, B. O. (2004) *Mplus Technical Appendices*. Los Angeles: Muthén and Muthén. (Available from `http://www.statmodel.com/mplus/techappen.pdf`.)
Muthén, L. K. and Muthén, B. O. (2003) *Mplus User's Guide*, 2nd edn. Los Angeles: Muthén and Muthén.
Pearl, J. (1995) Causal diagrams for empirical research. *Biometrika*, **82**, 669–710.
Pearl, J. (2001) Direct and indirect effects. In *Proc. 17th Conf. Uncertainty in Artificial Intelligence*, pp. 411–420. San Francisco: Morgan Kaufmann.
Power, C. and Hertzman, C. (1997) Social and biological pathways linking early life and adult disease. *Br. Med. Bull.*, **53**, 210–222.
Ross, C. E. and Wu, C. (1995) The links between education and health. *Am. Sociol. Rev.*, **60**, 719–745.
Sandana, R., Mathers, C. M., Lopez, A. D., Murray, C. J. L. and Iburg, K. M. (2002) Comparative analyses of more than 50 household surveys on health status. In *Summary Measures of Population Health, Concepts, Ethics, Measurement and Applications* (eds C. J. L. Murray, J. A. Salomon, C. Mathers and A. D. Lopez). Geneva: World Health Organization.
Singh-Manoux, A., Clarke, P. and Marmot, M. G. (2002) Multiple measures of socioeconomic position and psychosocial health: proximal and distal effects. *Int. J. Epidem.*, **31**, 1192–1199.
Sobel, M. E. (2000) Causal inference in the social sciences. *J. Am. Statist. Ass.*, **95**, 647–651.
Wadsworth, M. E., Butterworth, S. L., Hardy, R. J., Kuh, D. J., Richards, M., Langenberg, C., Hilder, W. S. and Connor, M. (2003) The life course prospective design: an example of the benefits and problems associated with study longevity. *Socl Sci. Med.*, **57**, 2193–2205.
White, I., Blane, D., Morris, J. and Mourouga, P. (1999) Educational attainment, deprivation-affluence and self-reported health in Britain: a cross-sectional study. *J. Epidem. Commty Hlth*, **53**, 535–541.
Winkleby, M. A., Jatulis, D. E., Frank, E. and Fortmann, S. P. (1992) Socioeconomic status and health: how education, income, and occupation contribute to risk factors for cardiovascular disease. *Am. J. Publ. Hlth*, **57**, 153–182.