HEALTH ECONOMICS
Health Econ. **21**(Suppl. 1): 129–150 (2012)
Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/hec.2815

MEDIATING ROLE OF EDUCATION AND LIFESTYLES IN THE RELATIONSHIP BETWEEN EARLY-LIFE CONDITIONS AND HEALTH: EVIDENCE FROM THE 1958 BRITISH COHORT

SANDY TUBEUF[a,*], FLORENCE JUSOT[b] and DAMIEN BRICARD[b]

[a]*AUHE, Leeds Institute of Health Sciences, University of Leeds, Leeds, UK*
[b]*PSL, Université Paris-Dauphine, LEDa-LEGOS, Paris, France*

ABSTRACT

The paper focuses on the long-term effects of early-life conditions with comparison to lifestyles and educational attainment on health status in a cohort of British people born in 1958. Using the longitudinal follow-up data at age 23, 33, 42 and 46, we build a dynamic model to investigate the influence of each determinant on health and the mediating role of education and lifestyles in the relationship between early-life conditions and later health. Direct and indirect effects of early-life conditions on adult health are explored using auxiliary linear regressions of education and lifestyles and panel Probit specifications of self-assessed health with random effects addressing individual unexplained heterogeneity. Our study shows that early-life conditions are important parameters for adult health accounting for almost 20% of explained health inequality when mediating effects are identified. The contribution of lifestyles reduces from 32% down to 25% when indirect effects of early-life conditions and education are distinguished. Noticeably, the absence of father at the time of birth and experience of financial hardships represent the lead factors for direct effects on health. The absence of obesity at 16 influences health both directly and indirectly working through lifestyles. Copyright © 2012 John Wiley & Sons, Ltd.

JEL Classification: D63; I12

KEY WORDS: cohort; early-life conditions; education; lifestyles; random effects

1. INTRODUCTION

Numerous literature references have agreed the important role played by current individual social characteristics, such as income, education level, wealth and social status (e.g. van Doorslaer and Koolman, 2004, Cutler *et al*., 2006, Lantz *et al*., 2010) in the explanation of health inequalities. More recently, several studies have also found early-life conditions as a relevant determinant of health inequalities with a large range of social background factors, such as low parental socioeconomic status (e.g. Currie and Stabile, 2003, Case *et al*., 2005, Lindeboom *et al*., 2009, Rosa-Dias, 2009, Jusot *et al*., 2010, Trannoy *et al*., 2010); family issues, such as living in a single parent family or experiencing marital discord (Case and Katz, 1991, Francesconi *et al*., 2010); parents' health status (Trannoy *et al*., 2010); or health-risk lifestyles (Anda *et al*., 2002, Göhlmann *et al*., 2010, Jusot *et al*., 2010). However, the importance of lifestyles in the magnitude of health inequalities is less clear. Whereas epidemiological literature concluded until recently that lifestyles make a relatively minor contribution to the social gradient in health (Van Oort *et al*., 2005, Khang *et al*., 2009, Skalicka *et al*., 2009, Lantz *et al*., 2010), health economists have shown that differences in lifestyles can

*Correspondence to: Sandy Tubeuf, Academic Unit of Health Economics, Leeds Institute of Health Sciences, University of Leeds, Charles Thackrah Building, 101 Clarendon Road, Leeds LS2 9LJ (UK). E-mail: s.tubeuf@leeds.ac.uk

explain a relevant part of health and mortality inequalities (Contoyannis and Jones, 2004, Häkkinen *et al*., 2006, Balia and Jones, 2008), and a few recent epidemiological studies (Strand and Tverdal, 2004, Laaksonen *et al*., 2008, Menvielle *et al*., 2009, Stringhini *et al*., 2010) have also confirmed that the impact of lifestyles on health and mortality disparities would be larger than it was previously estimated, particularly if lifestyles are observed longitudinally.

The issue at stake is that early-life conditions, education, and lifestyles cannot be considered as independent.

Several studies provide evidence on the transmission of socioeconomic status over different generations such as social class, education level or income level (e.g. Marmot *et al*., 2001, Case *et al*., 2005, Trannoy *et al*., 2010). Moreover, parents' lifestyles and social status as well as early-life conditions would also be associated with health-related behaviours in later life such as smoking (Rosa-Dias, 2009; Göhlmann *et al*., 2010; Francesconi *et al*., 2010), alcohol consumption (Anda *et al*., 2002); obesity (Power *et al*., 2005; Laitinen *et al*., 2001; von Hinke Kessler Scholder, 2008). In the sense that early-life conditions such as parents' lifestyles, socioeconomic status as well as childhood economic conditions are individual characteristics from childhood, one would agree that they are causal factors of adult education and lifestyles. The sense of causality between education and lifestyles in adulthood is less straightforward. Poor lifestyles in adolescence such as heavy drinking, cannabis and other drug consumption have been found to affect achieved education (Koivusilta *et al*., 1999; Staff *et al*., 2008; Van Ours and Williams, 2009); but studies uncovering mechanisms through which education affects lifestyles are much more numerous. In particular, education has been largely found to affect smoking decisions (Kenkel, 1991; Grimard and Parent, 2007; de Walque, 2007; Reinhold and Jürges, 2010; Etilé and Jones, 2011; Kemptner *et al*., 2011; Kjellsson *et al*., 2011; Frisvold and Golberstein, 2011; Jürges *et al*., 2011). The causal effect of education has also been found on obesity (Webbink *et al*., 2010; Reinhold and Jürges, 2010; Frisvold and Golberstein, 2011; Kemptner *et al*., 2011) and exercise (Kenkel, 1991, Park and Kang, 2008).

Relying on the literature, we consider an empirical analysis of the interplay between these three broad determinants (Figure 1). The objective of this study is to explore the long-term effects of social and health-related early-life conditions, education, and lifestyles on health and to evaluate their respective contribution to the magnitude of health inequalities. Based on a dynamic model of health status over the life cycle, our empirical analysis aims to investigate the effect of each determinant on overall health inequality and determines whether early-life conditions influence health directly or indirectly, that is via affecting education and lifestyles.

Our findings provide new elements on the determinants of health inequalities, which are relevant for policymakers and that remained to be empirically assessed. First, the role of early-life conditions is explored in direct and indirect terms with a larger set of indicators than previous analyses, including parental social and health conditions in addition to the individual's initial health status. Second, this research analyses the evolution of unhealthy lifestyles, their changes over an extended period and their association with health status. Finally, the longitudinal dimension of those data allows using dynamic panel analysis to control for unexplained individual heterogeneity and explain impact of past health status.

The structure of the paper is as follows. Section 2 describes the model that is empirically tested. Section 3 presents the National Child Development Study (NCDS) data and the variables of interest. Section 4 presents the empirical results, and section 5 concludes.

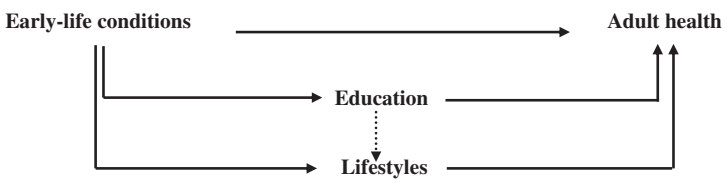

Figure 1. Early-life conditions, education, lifestyles and later-life health status

DOI: 10.1002/hec

2. THE MODEL

2.1. General health production function

In contrast with Jusot *et al*. (2010), who focused on a reduced-form model of childhood circumstances and lifestyles, we use a full model specification including individual educational attainment. Our approach also differs from Contoyannis and Jones (2004), Häkkinen *et al*. (2006), and Balia and Jones (2008), as our health production function includes early-life conditions as a potential determinant for health in addition to education and lifestyles. Furthermore, we built a dynamic model of health using longitudinal data.

The individual health status H can be written using the following health production function:

$$H = f(C, E, L, D, e) \quad \text{(Eq. 1)}$$

The vector of early-life conditions C consists of a set of variables beyond individual control, which may be related to health status. The literature on health determinants suggests an influence of childhood conditions and family background on health status in adulthood, (see, for example, Currie and Stabile, 2003, Case *et al*., 2005, Rosa-Dias, 2009, Lindeboom *et al*., 2009, Trannoy *et al*., 2010). Moreover, initial health, such as birth weight and health problems during childhood and adolescence, also significantly influences health in adulthood, and the most adverse health risks in adulthood tend to be experienced by people in poor health in childhood and adolescence (Moser *et al*., 2003, Case *et al*., 2005). The vector E represents individual's education level and is not a time-variant variable. Researchers in many countries have found a relevant and persistent association between education and health as measured by various health measures (Grossman, 2006), and several recent studies have found a causal effect of education on health status (Oreopoulos, 2006; Silles, 2009; Kemptner *et al*., 2011). The vector of health-related behaviours L captures individual decisions to invest in health capital, such as lifestyles (Contoyannis and Jones, 2004, Häkkinen *et al*., 2006, Balia and Jones, 2008, Rosa-Dias, 2009, Jusot *et al*., 2010). The vector D represents demographic characteristics, which are biological determinants of health status, only captured by gender in cohort data. Finally, the residual term e represents unobserved heterogeneity related to other random factors, which cannot be captured by observed determinants.

More concretely, let us assume that health of individual i at wave t is measured by a continuous latent variable H_{it}^* which is proxied using a binary variable H_{it} as follows:

$$H_{it}^* \geq 0 \text{ when } H_{it} = 1$$

$$H_{it}^* < 0 \text{ when } H_{it} = 0$$

The general health production model can be written as follows:

$$H_{it}^* = a_1 C_i + a_2 D_i + b_1 E_i + c L_{it-1} + u_i + v_{it}$$

with $i = 1, \ldots, N$ and $t = 1, \ldots, T_i$ (Eq. 2)

We include lagged values of lifestyles into the model L_{it-1} as past lifestyles are more likely to be important for health status than just acquired lifestyles. The time-variant individual specific error term is captured by v_{it}, which is assumed to be normally distributed and uncorrelated across individuals and waves. The individual time invariant unobserved effect is captured by u_i; it captures unobserved individual characteristics, such as genetics and personality traits. We first estimate a static model with random effects (model 1a) assuming that the errors are independent over time and uncorrelated with the explanatory variables. This model provides us with base estimates, with which we can compare results from models that incorporate unobserved heterogeneity and state dependence.

Model 1a does not allow us to address several important issues with relevant impact on the health determinants. First, we do not know whether the model variables appropriately account for any unobserved individual characteristics that also influence time-variant variables. Especially, if the past lifestyles are correlated with u_i, we would expect to overestimate the effect of lifestyles in model 1a. Second, early-life conditions, education and past

DOI: 10.1002/hec

lifestyles may affect current health directly but also indirectly, namely, through affecting past health. If this is true, we would expect past health to influence current health and the direct effects of early-life conditions, education and past lifestyles on current health to weaken or even disappear. Finally, the initial health state is likely to be not randomly assigned to the individual. To address the first issue, we use a random effect Probit specification allowing u_i and v_{it} to be correlated and introduce lifestyles averaged over time $\bar{L}_i$ as a set of controls for unobserved heterogeneity (Mundlak, 1978). We now estimate the effects of changing lagged lifestyles on health but holding the average fixed[1] in model 1b. Although this model addresses part of the problem of unobserved heterogeneity, a dynamic model of health that incorporates both past health and unobserved effects is required to address the remaining issues. The inclusion of a rich set of early-life conditions in the model can be interpreted as a particular specification of the individual component. A well-explained vector strongly contributes to the reduction of the correlation between individual effects and initial conditions as it minimises unobserved time-invariant characteristics affecting individual outcomes at each point of time. Nevertheless, we need to account the potential endogeneity bias related to the respective correlations between early-life conditions, lifestyles and education with past health, which can be ruled out using a dynamic specification and introducing past health H_{it-1} into the health production function (model 1c). The introduction of past health status in our empirical model allows us to capture the state dependence in health reports and strongly reduces the impact of individual heterogeneity. In a dynamic context, initial health H_{i0} is likely to be correlated with unobserved heterogeneity u_i affecting H_{it}, and if H_{i0} is considered exogenous, this will lead to inconsistent estimators. We follow the alternative approach suggested by Wooldridge[2] (2002), which requires to specify the distribution of u_i given H_{i0} and other exogenous variables and, so, include at least the first value of the independent variable, H_{i0}.

The ultimate latent health model that we estimate can be written as follows:

$$H_{it}^{*} = a_1 C_i + a_2 D_i + b_1 E_i + c_1 L_{it-1} + c_2 \bar{L}_i + d_1 H_{it-1} + d_2 H_{i0} + u_i + v_{it}$$

with $i = 1, \ldots, N$ and $t = 1, \ldots, T_i$ (Eq. 3)

In this model, the three broad determinants of health are, in some aspects, considered as independent; therefore, we complement this primary specification with a mediating specification incorporating interdependent relationships between early-life conditions, education and lifestyles.

2.2. Mediating effects identification

The mediating specification aims to identify whether explanatory variables influence health directly or indirectly, that is by affecting or being affected by another explanatory variable.

Let us first consider a more general case where individual health status H is defined according to a set of variables C, such as her early-life conditions, and a set of variables M, such as her education or lifestyles.

$$H = F(C, M, \varepsilon) \quad \text{(Eq. 4)}$$

We consider that M potentially mediates the relationship between C and H. For example, mother's educational qualification may affect adult health through an effect on individual's educational attainment as

[1]Lifestyles could therefore be regarded both as a measure of lifestyles shocks on health via the past lifestyle variables and as a measure of long-term or 'permanent' lifestyles on health via the average lifestyle. Nevertheless, from our point of view, the follow-up of lifestyles being limited to four points of time and to the use of binary lifestyle variables does not justify to interpret the effects of lifestyles on health in terms of permanent and transitory effects.

[2]Two other methods to address initial conditions problems could have been considered: Heckman (1981) and Orme (2001). The former suggests approximating the reduced form $f(H_{i0}|x_i, u_i)$ and then specifying $f(u_i|x_i)$; $f(H_{i0}, \ldots, H_{iT}|x_i)$ is then given by integrating out u_i (where x_i includes all the regressors). The two main difficulties with this method are specifying the distribution of initial health and computing time. As for Orme (2001), he suggested a two-step bias corrected procedure that is locally valid when the correlation between H_{i0} and H_{it} is approximated to zero. A couple of recent works compared the relative performance of the three methods. Whereas Miranda (2007) concluded that the Heckman method delivers estimators that are hardly subject to bias and that are estimated with high precision, Arulampalam and Stewart (2009) concluded that none of the three estimators dominate the other two in all cases. Moreover, the authors found that it is advantageous to allow for correlated random effects using the approach of Mundlak (1978).

DOI: 10.1002/hec

exhibited in the *pathway model* that has been well studied in both economic and epidemiological studies (Marmot *et al*., 2001, Case *et al*., 2005, Trannoy *et al*., 2010). We aim to evaluate the full effect of C on H such as the following:

$$H = \alpha_1 C + \varepsilon_1 \quad \text{(Eq. 5a)}$$

However, as we are in a full model specification, we cannot ignore the role played by M on H:

$$H = \alpha_2 C + \beta M + \varepsilon_2 \quad \text{(Eq. 5b)}$$

The total effect of C on H is measured by α_1, whereas the direct effect of C on H is measured by α_2. The difference between α_1 and α_2 represents the mediating effects of C on H that work through M. Moreover, the mediating effects can be written using the following auxiliary equation, where φ captures the effect of C on M:

$$M = \varphi C + \omega \quad \text{(Eq. 6)}$$

Using a linear model to estimate the relationship between M and C, we can rewrite (Eq. 5b) as follows:

$$H = \alpha_2 C + \beta(\varphi C + \omega) + \varepsilon_2$$

$$H = (\alpha_2 + \beta.\varphi)C + \beta.\omega + \varepsilon_2 \quad \text{(Eq. 5c)}$$

where $\beta.\varphi$ represents the mediating effect, namely, the indirect effect of C on H working through M. Prior to the estimation of Eq. 5c, estimated residuals $\hat{\omega}$ must be estimated from the auxiliary equations (Eq. 6).

Following Bernt Karlson *et al*. (2012) and Bernt Karlson and Holm (2011), we can express the respective direct, mediating and total effects of C on H as follows[3]:

Direct effects	α_2
Mediating effects	$\beta.\varphi$
Total effects	$\alpha_1 = \alpha_2 + \beta.\varphi$

Let us now consider our present study, using Figure 1, the set of variables C represents early-life conditions, and the set of variables M is both education and lifestyles. In this context, the mediating specification aims to describe whether early-life conditions influence health directly or indirectly, that is via affecting education and lifestyles.

In addition, the dashed arrow in Figure 1 suggests that the set of variables C could represent both early-life conditions and education and the set of variables M be lifestyles only. Relying on numerous studies showing the causal impact of education on lifestyles, it is likely that a share of the impact of education on adult health is mediated via lifestyles. Our empirical analysis thus distinguishes two layers of mediating effects: mediating effects of early-life conditions on health working through education and lifestyles (mediating specification 1) and mediating effects of education on health working through lifestyles (mediating specification 2).

The second specification ignores potential reverse causality link of lifestyles on education achievement. As mentioned in the Introduction, our motivation for ignoring reverse causality is based on the enhanced literature on the causal effect of education on lifestyles. Nevertheless, we also use the panel dataset thoughtfully to address econometrically this matter.

The two different mediating specifications will be tested and compared. In concrete terms, Eq. 5b corresponds to the general health production function (Eq. 3, model 1c presented earlier), whereas the mediating specification is a two-step estimation based on auxiliary equations and then the estimation of the health production function described in (Eq. 5c). The sets of auxiliary equations being estimated can be written as follows:

[3]The introduction of the estimated residuals from the auxiliary equations into the main equation ensures that the total effect is equal to the sum of the direct and mediated effects (see Bernt Karlson *et al*., 2012)

DOI: 10.1002/hec

$$E_i = \varphi_1 C_i + \varphi_2 D_i + e_i \quad \text{(Eq. 6a)}$$

$$L_{1,it} = \varphi_1^a C_i + \varphi_2^a D_i + l_{1,i} \quad \text{(Eq. 6b, mediating specification 1)}$$

$$L_{2,it} = \varphi_1^a C_i + \varphi_2^a D_i + \varphi_3^a E_i + l_{2,i} \quad \text{(Eq. 6b, mediating specification 2)}$$

$$\overline{L_{1,i}} = \varphi_1^b C_i + \varphi_2^b D_i + \overline{l_{1,i}} \quad \text{(Eq. 6c, mediating specification 1)}$$

$$\overline{L_{2,i}} = \varphi_1^b C_i + \varphi_2^b D_i + \varphi_3^b E_i + \overline{l_{2,i}} \quad \text{(Eq. 6c, mediating specification 2)}$$

If we replace the respective auxiliary equations of the two mediating specifications into the main equation (Equation 3, model 1c), the health production function in the mediating specification becomes[4] the following:

$$H_{it}^* = (a_1 + \psi_1)C_i + (a_2 + \psi_2)D_i + b_1\hat{e}_i + c_1\hat{l}_{1,it} + c_2\hat{\bar{l}}_{1,i} + d_1 H_{it-1} + d_2 H_{i0} + u_i + v_{it}$$

with $i = 1, \ldots, N$ and $t = 1, \ldots, T_i$ (Eq. 7, mediating specification 1)

$$H_{it}^* = (a_1 + \psi_1)C_i + (a_2 + \psi_2)D_i + \psi_3 E_i + b_1\hat{e}_i + c_1\hat{l}_{2,it} + c_2\hat{\bar{l}}_{2,i} + d_1 H_{it-1} + d_2 H_{i0} + u_i + v_{it}$$

with $i = 1, \ldots, N$ and $t = 1, \ldots, T_i$ (Eq. 7, mediating specification 2)

where, $\hat{e}_i$, $\hat{l}_{1,it}$, $\hat{\bar{l}}_{1,i}$ (respectively $\hat{l}_{2,it}$, $\hat{\bar{l}}_{2,i}$ in mediating specification 2) represent the estimated residuals in each auxiliary equation and can be written as follows:

$$\hat{e}_i = E(e_i|C_i, D_i) \quad \text{(Eq. 7a, mediating specification 1)}$$

$$\hat{l}_{1,it} = E\left(l_{1,i}|C_i, D_i\right) \quad \text{(Eq. 7b, mediating specification 1)}$$

and

$$\hat{l}_{2,it} = E\left(l_{2,i}|C_i, D_i, E_i\right) \quad \text{(Eq. 7b, mediating specification 2)}$$

$$\hat{\bar{l}}_{1,i} = E\left(\bar{l}_{1,i}|C_i, D_i\right) \quad \text{(Eq. 7c, mediating specification 1)}$$

and

$$\hat{\bar{l}}_{2,i} = E\left(\bar{l}_{2,i}|C_i, D_i, E_i\right) \quad \text{(Eq. 7c, mediating specification 2)}$$

These estimated residuals[5] are estimated as linear probability models for time-invariant outcomes and pooled linear probability models otherwise. The estimated residuals are then introduced in the health equation (Eq. 7) in replacement of the actual explanatory variables as carried out in (Eq. 5). In (Eq. 7), we can express the respective direct, mediating (indirect) and total effects of early-life conditions and education on health as follows:

Direct effects of early-life conditions on health	a_1
Total mediating effects of early-life conditions on health	$\psi_1 = b_1\varphi_1 + c_1\varphi_1^a + c_2\varphi_1^b$
Mediating effects via education	$b_1\varphi_1$
Mediating effects via lifestyles	$c_1\varphi_1^a + c_2\varphi_1^b$
Mediating effects of education on health via lifestyles	$\psi_3 = c_1\varphi_3^a + c_2\varphi_3^b$

[4]Let us consider the mediating specification 1, if we replace (Eq. 6a), (Eq. 6b) and (Eq. 6c) in (Eq. 3, model 1c), we obtain $H_{it}^* = a_1 C_i + a_2 D_i + b_1(\varphi_1 C_i + \varphi_2 D_i + \hat{e}_i) + c_1\left(\varphi_1^a C_i + \varphi_2^a D_i + \hat{l}_i\right) + c_2\left(\varphi_1^b C_i + \varphi_2^b D_i + \hat{\bar{l}}_i\right) + d_1 H_{it-1} + d_2 H_{i0} + u_i + v_{it}$ therefore, we have $H_{it}^* = \left(a_1 + b_1\varphi_1 + c_1\varphi_1^a + c_2\varphi_1^b\right)C_i + \left(a_2 + b_1\varphi_2 + c_1\varphi_2^a + c_2\varphi_2^b\right)D_i + b_1 e_i + c_1 l_i + c_2\bar{l}i + d_1 H_{it-1} + d_2 H_{i0} + u_i + v_{it}$. If we assume $\psi_1 = b_1\varphi_1 + c_1\varphi_1^a + c_2\varphi_1^b$, $\psi_2 = b_1\varphi_2 + c_1\varphi_2^a + c_2\varphi_2^b$, then we obtain (Eq. 7) for the mediating specification 1.

[5]For binary outcomes, one estimated residual will be predicted from the OLS estimation.

DOI: 10.1002/hec

2.3. Health determinants decomposition

The second part of our empirical analysis inquires to which extent the account of mediating effects influences the contribution of each determinant to health disparities. Shorrocks (1982) showed that if we are interested in an absolute measure of inequality, the variance is a good index, and its natural decomposition presents the desired properties. The alternative model specifications of the health production function we considered in sections 2.1 and 2.2 are based on strictly identical regressors, and so, they have the same variance. The standard errors of the variance of both models is estimated using bootstrap method (see statistical inference sub-section).

In a linear case, the share of variance explained for example by early-life conditions C_i simply consists in the share of the R^2 of the model, which is explained by C_i. In a nonlinear context, it is not straightforward as H^*_{it} can only be measured as a prediction, and ω_i and ε_{it} are defined as independent of the set of K explicative variables. A variance estimated from the data is attributed to the time-invariant individual error term ω_i, whereas the time-variant individual error term ε_{it} has a variance normalised to be equal to 1 in the case of a Probit model. We use the pseudo R^2 proposed by McKelvey and Zavoina (1975) to measure the share of variance explained by the variable X^k having an associated coefficient η^k, which is based on predictions of the latent endogenous variables:

$$\hat{H}^*_{it} = \sum \eta^k X^k_{it} \qquad \text{(Eq. 8)}$$

Assuming that the variance of the error term u_i and v_{it} follows a normal distribution in this random effect Probit model, we can write

$$R^2 = \frac{V(\hat{H}^*)}{V\left(\hat{H}^*\right) + \sigma_u + 1} \qquad \text{(Eq. 9)}$$

Given the longitudinal data, the variance of the latent health variable can either be decomposed directly using the explained variance of the latent health variable, as measured by the pseudo-R^2 based on all the waves. The share of inequality associated to the variable X^k can thus be written as

$$I\left(X^k\right) = \frac{cov\left(\hat{H}^*, \eta^k X^k\right)}{V\left(\hat{H}^*\right)} = \frac{cov\left(\hat{H}^*, \eta^k X^k\right)}{\sum_{k=1}^{K} cov\left(\hat{H}^*, \eta^k X^k\right)} \qquad \text{(Eq. 10)}$$

2.4. Statistical inference

All our random effect Probit models have been bootstrapped to test the robustness of our estimated coefficients using 300 replications, and this provides us with standard errors.

This is particularly relevant for the two-step estimation needed for the mediating specification, which is likely to introduce uncertainty because estimated residuals from the auxiliary equations are introduced in the main health equation. From those, we produce standard errors for the direct and total effects. The significance level of the estimated indirect effects is tested using the statistical test generated by Kohler *et al.* (2011) and available as the *khb* command in Stata.

The health inequality decomposition results are also tested using the same statistical inference where the whole analysis (auxiliary equations, health equation and the decomposition) is replicated 300 times. This provides us with standard errors and confidence intervals based on the percentile method for the health inequality decomposition by sources.

3. THE NATIONAL CHILD DEVELOPMENT STUDY

The National Child Development Study (NCDS) is a continuing, multi-disciplinary longitudinal study, which focuses on all the people born in one week in March 1958 in England, Scotland and Wales. Information was

DOI: 10.1002/hec

gathered from almost 17 500 babies. Following the initial birth survey in 1958, there have been seven attempts to trace all members of the birth cohort to monitor their physical, educational, social and economic development. These were carried out in 1965, 1969, 1974, 1981, 1991, 1999/2000 and 2004. For the birth survey, information was obtained from the mother and from medical records by the midwife. For the purpose of the first three NCDS surveys, information was obtained from parents, head teachers and class teachers, the schools health service and the subjects themselves (who completed tests of ability and, latterly, questionnaires). In the 1981 and later surveys, information was gathered by professional survey research interviewers. In 1981, information was obtained from cohort members and from the 1971 and 1981 Censuses. In the 1991 survey, there was a professional interview with the cohort member along with self-completion questionnaires from NCDS subjects and husbands, wives and cohabiters. For the 1999–2000 sweeps, information was obtained from cohort members by interviewer and self-completion using CAPI. The 2004 survey was administered by telephone.

3.1. The sample

For the purpose of our study, we focus on the four last sweeps of the cohort ($t = 0, \ldots, 3$) to have repeated measures of both lifestyles and health status as an adult. Data collected before age 23 are used to inform individual early-life conditions (see in Appendix Table A.I). We have excluded cohort members who missed at least one of the four first sweeps to ensure a description of childhood conditions with a limited nonresponse. The balanced sample for which individuals have fully informed health status and lifestyles in all the sweeps 4 to 7 contains 4,480 individuals, whereas the unbalanced sample contains between 5,900 and 7,900 individuals. A description of the distribution of relevant variables in the balanced sample at $t = 0$ is available in appendix (see Table A.II)

3.2. Health variable

The NCDS includes only one repeated measure of the respondent's health in the cohort, namely, self-assessed health (SAH). Respondents are asked to rate their own health on a four-point or five-point categorical scale ranging from poor (sweeps 4, 5 and 6) or very poor (wave 7) to excellent health status. Given the changes in scale in the variable over the different waves, we use SAH as a binary variable[6] which takes the value one if the individual rates her health as good health or higher, and zero if she rates her health less than 'good'. Self-assessed health has been shown to be a good predictor of mortality, morbidity and subsequent use of health care (Idler and Benyamini, 1997). The distribution of health status in the balanced sample shows the age effect on health status over the life cycle (see Table A.III). Whereas good health represents 92.7% of respondents at 23 years old, the proportion of respondents reporting a good health declines to 78.3% at 46 years old. Between the first three sweeps, the mean is declining by a constant rate of four percentage points. There is a break with a decrease of six percentage points between the two last sweeps despite they are separated by 4 years only. This difference could be explained by an increasing effect of ageing on health when the cohort member enters her 40s. This shift could also come from the change in the categorical scale of self-assessed health between sweep 6 and sweep 7, and this latter issue is minimised by the dichotomisation of health.

3.3. Education

The NCDS provides several current social characteristics, such as income or professional status, but to prevent us from endogeneity that may be observed between social characteristics with either health and lifestyles, we only consider educational qualification in the analysis. We particularly focus on a binary measurement of education based on whether the cohort member has at least being awarded an O-level or not. O-level is a subject-based qualification conferred as part of the general certificate of education, and it is designed for 14- to 16–year-old

[6]Dichotomisation was also required as the necessary condition on the proportional odds assumption for ordinal Probit models turns out not to be valid.

DOI: 10.1002/hec

pupils. Therefore, educational qualification is measured during adolescence, and so, as described in Figure 1, it is expected that lifestyles in adulthood will be affected by educational level in this context and not the reverse causal link. About one fifth of respondents have an educational qualification strictly lower than O-level.

3.4. Lifestyles variables

The NCDS includes a longitudinal follow-up of lifestyles and health records at age 23, age 33, age 42 and age 46. We consider five lifestyles binary variables (presented in Table A.IV). Exercising indicates whether the cohort member is regularly doing exercise or sports; it equals one if the cohort member report exercising and zero otherwise. Nonsmoking informs whether the cohort member is a current smoker at the time of the wave; it equals one if she does not currently smoke and zero otherwise. Drinking prudently is a gender-specific indicator based on the number of units of alcohol drinks taken the week before the interview. Male subjects are considered to drink prudently if they drank between 0 and 21 units of alcohol, whereas it is between 0 and 14 units a week for female subjects (Working Party of the Royal College of Physicians UK, 2001, Balia and Jones, 2008). The binary variable takes the value one if the respondent drinks prudently and zero otherwise. The absence of obesity is the fourth lifestyle that we consider. Obesity may appear as a genetic outcome of health and not a pure lifestyle. Given that we can control the genetic and the family-transmitted effect on obesity using the respondent's obesity status when she was 16, the absence of obesity will thus capture aggregated effects of lifestyles. Absence of obesity is constructed using the reported height and weight and calculating individuals' body mass index (BMI[7]). The absence of obesity is a binary variable taking the value one if the cohort member's BMI is strictly lower than 30 and zero otherwise. Furthermore, we consider the frequency of eating fruits or vegetables as processed, raw, cooked, canned or in salads. Diet information was only collected at age 33, 42 and 46 in the cohort. Therefore, we assume that diet at age 23 was the same as that at age 33 and consider a binary variable taking the value one if the respondent reports that she eats at least one fruit or one vegetable once per week and zero otherwise.

3.5. Early-life conditions

The vector of early-life conditions that we consider has three main types of variables: social conditions in childhood, parents' health and health-related behaviours, and child and adolescence health. Social conditions in childhood include the father's social class at the time of birth, the father and the mother's education levels and parental reports of financial hardships when the cohort member was 16. Father's social class is described in three large categories: a top class (I/II) including professional and managerial or technical workers, a middle class (III) including skilled workers and armed services and a low class (IV/V) including partly skilled and unskilled workers and a fourth category is added if the mother reported no male figure in the household at the time of birth. Parental education consists of a two-category variable: parents who dropped out from school before or at the minimum age (14–15 years depending on the year of birth) and parents who were still at school after this age. Parents' health is measured by parental report of chronic illness when the cohort member was 16 years old. Regarding parents' health-related lifestyles, we used a smoking indicator for each parent taking the value one if reported to be a smoker. Respondent's health in childhood and adolescence are used as control variables but also as achievement variables because they may represent health-related difficulties from the living environment during childhood. We use the same approach as Case *et al*. (2005) who considered the report of at least one chronic condition at 16 as well as a birth weight below 2.5 kg as health indicators before adulthood using the same dataset. Furthermore, we include obesity at 16 years old. We have computed BMI using medical assessment of height and weight and evaluated obesity level using gender-specific thresholds values found to be a good predictor of obesity at 18 (Lahti-Koski and Gill, 2004).

[7]BMI in kg/m^2 = weight/height².

DOI: 10.1002/hec

4. RESULTS

4.1. Random effect dynamic panel Probit results

The results of the random effect panel Probit of the general specification are presented in Table I. Three different models are reported: model 1a and model 1b are static models with random effect, with model 1b including the average individual lifestyles over the studied period, whereas model 1c is a dynamic random effects Probit model.

The results show that several early-life conditions have a statistically significant effect on the probability to report good health regardless of the model. Individuals whose father belonged to the lowest social class, namely, partly skilled and unskilled workers as well as individuals who had no father at the time of their birth are significantly less likely to report good health. Similarly, the experience of financial hardship during childhood has a significant and negative effect on reports of good health. The mother's education level is also found as a statistically significant determinant of poor health reports, whereas the father's education is not significant in any of the models. This may be explained by father's social class being significant and, so, absorbing the effect of father's education on descendant's health. Unlike mother's illness, father's illness significantly reduces the probability to report good adult health. Mother's smoking behaviour appears to be significant for descendant's report of good health, but the significance level weakens in the dynamic model. Individual education level is also found statistically and significantly associated with health: low educational attainment is negatively associated with the report of a good health. Regarding lifestyles variables, the five lagged lifestyles are significantly associated with reports of good health in model 1a (at the 10% level for drinking prudently). However, as soon as average lifestyles are added to the model, the lagged lifestyles are not significantly associated with health anymore. The average behaviours in exercise, absence of obesity, consumption of fruits and vegetables and, to a lesser extent, the absence of smoking are found strongly and significantly associated with reports of good health, whereas the average behaviour towards drinking is significant at the 10% level only.

This first table of results emphasises the relevance of a dynamic model; past health status and initial health in model 1c are significantly associated with reports of good health and the share of individual unexplained heterogeneity addressed by the model equals 37% of the unexplained heterogeneity. Furthermore, when past and initial health variables are introduced, a reduction in the magnitude of the model's estimated coefficients is observed.

4.2. Auxiliary equations estimations

Prior to the mediating specification, the auxiliary equations are estimated. The results are presented in Appendix (Table A.V). Education is estimated as a linear regression model of the probability to report an educational qualification lower than O-level. Having an educational qualification lower than O-level is actually found positively and significantly associated with father's low social class and absence of a male head at the time of birth, experience of financial hardship, both parents' having left school before or at the minimum legal age and both parents' being smokers, presence of a chronic disease at 16 years old, obesity at 16 and low birth weight. The five lifestyles are estimated as OLS models on the pooled sample. Exercising is significantly and negatively associated with reports of chronic disease at 16, father's low social class, both parents' low education and experience of financial hardship. When educational attainment is controlled (Table A.V, model b), educational qualifications is found significantly and negatively associated with regular exercise. Moreover, it weakens the significant effect of financial hardships, father's social class and parental education. With regard to smoking behaviour, the absence of smoking is strongly and negatively associated with both parents' smoking, financial hardship at 16, father's low social class and the absence of male head at the time of birth. When introduced in the model, educational qualification lower than O-level appears to be negatively correlated with the absence of smoking (Table A.V, model b). Moreover, the inclusion of individual education absorbs the effect of father's social class and weakens the impact of financial hardship and mother's smoking. Drinking prudently is strongly higher among female and found negatively and significantly correlated with father's smoking. Noticeably, having a father who was in lower social class and a low birth weight appear to be positively associated with prudent drinking.

DOI: 10.1002/hec

Table I. Random effect Probit models results with estimated coefficients (with bootstrapped standard errors)

	Model 1a *Static model*			Model 1b *Mundlak specification*			Model 1c *Dynamic model*		
Variables	**Coef.**		**SE**	**Coef.**		**SE**	**Coef.**		**SE**
Gender (male)	0.115	**	0.050	0.094	*	0.052	0.070		0.047
Early-life conditions									
Fathers' social class (Ref.: I and II - Professional and managerial/technical)									
III - Skilled	−0.103		0.076	−0.087		0.076	−0.064		0.066
IV and V - Partly skilled and unskilled	−0.275	***	0.091	−0.255	***	0.092	−0.206	***	0.076
No male head	−0.510	***	0.141	−0.492	***	0.143	−0.383	***	0.109
Financial hardship (Ref.: None)									
Yes	−0.347	***	0.090	−0.340	***	0.091	−0.253	***	0.070
Non response	0.057		0.165	0.080		0.166	0.102		0.149
Father's education (Ref.: beyond the min age)									
Before or at the min age	−0.092		0.071	−0.069		0.072	−0.046		0.061
Mother's education (Ref.: beyond the min age)									
Before or at the min age	−0.172	**	0.068	−0.152	**	0.069	−0.143	**	0.061
Parental illness (Ref.: None)									
Father's illness	−0.229	**	0.091	−0.218	**	0.092	−0.165	**	0.078
Mother's illness	−0.158		0.108	−0.142		0.109	−0.120		0.092
Parental smoking (Ref.: None)									
Father's smoking	0.087		0.054	0.113	**	0.054	0.076	*	0.048
Non response	0.024		0.105	0.023		0.106	−0.009		0.093
Mother's smoking	−0.142	***	0.052	−0.127	**	0.053	−0.077	*	0.044
Non response	−0.113		0.138	−0.108		0.139	−0.064		0.126
Chronic condition at 16 (Ref.: None)									
Yes	−0.114		0.075	−0.090		0.076	−0.016		0.065
Non response	0.190		0.154	0.157		0.155	0.113		0.131
Low birth weight	−0.124		0.110	−0.131		0.111	−0.074		0.089
Obesity at 16 (Ref.: Yes)									
No	0.062		0.203	−0.281		0.213	−0.316	*	0.194
Non response	−0.197		0.138	−0.171		0.138	−0.153		0.118
Educational qualification									
Strictly lower than O-level	−0.298	***	0.062	−0.191	***	0.063	−0.169	***	0.054
Lifestyles									
Lagged lifestyles									
Exercising once in the last 4 weeks	0.159	***	0.042	−0.033		0.049	−0.045		0.046
Non smoking	0.308	***	0.047	0.080		0.074	0.070		0.074
Drinking prudently	0.101	*	0.059	0.031		0.072	0.033		0.071
Non obese	0.361	***	0.065	−0.027		0.085	−0.052		0.087
Consumption of fruit/veg. once per week	0.324	***	0.067	0.095		0.091	0.072		0.096
Mean lifestyles									
Exercising once in the last 4 weeks				0.689	***	0.098	0.524	***	0.093
Non smoking				0.326	***	0.097	0.205	**	0.090
Drinking prudently				0.223	*	0.133	0.211	*	0.119
Non obese				0.924	***	0.141	0.771	***	0.129
Consumption of fruit/veg. once per week				0.422	***	0.135	0.318	***	0.123
Lagged health status							0.306	***	0.076
Initial conditions (health status at 23)							0.997	***	0.101
Time dummies (Ref.: t = 3)									
t = 1	0.670	***	0.045	0.666	***	0.047	0.578	***	0.049
t = 2	0.347	***	0.040	0.384	***	0.040	0.340	***	0.038
$V\left(\hat{H}^*\right)$	0.247			0.342			0.367		
σ_ω	1.110			1.113			0.636		
$\rho^{\#}$	0.526			0.527			0.389		
R^2 (McKelvey and Zavoina)	0.105			0.139			0.183		

[#]Share of individual unexplained heterogeneity measured by the share of σ_ω in the total unexplained variance ($\sigma_\omega + 1$).
Significance levels: ***1%, **5%, *10%.

DOI: 10.1002/hec

Educational qualification is found to be negatively associated with drinking prudently when education is introduced in the auxiliary equation (Table A.V, model b). The absence of obesity at 16 is statistically associated with non obesity in adulthood; in addition, mother's smoking, father's SES and mother's low education are found statistically significant for the reduction of the absence of obesity. When education is included within the auxiliary equation (Table A.V, model b), it appears to be significantly and negatively associated with the absence of obesity. Moreover, the introduction of education erases the significant effect of parents' education, which was previously observed. Finally, eating fruits and vegetables is found negatively and significantly correlated with father's qualification, mother's education, father's smoking (Table A.V, models a and b) and respondent's education level (Table A.V, model b).

4.3. Random effect dynamic panel Probit results: mediating specification

The results of the mediating specifications of the health production function are presented in Table II. They have been obtained by replacing actual variables of education and lifestyles by the estimated residual terms of the different auxiliary equations whose results were described in the previous section. The two mediating specifications are presented.

Noticeably, the estimated coefficients associated to education and lifestyles in mediating specification 1 and to lifestyles only in mediating specification 2 are strictly identical with the estimated coefficients in model 1c (Table I) as expected when comparing Eq. 3 with respectively Eq. 7 mediating specification 1 and Eq. 7 mediating specification 2. The results of the mediating specification permit confirming the existence of indirect effect of early-life conditions variables on health over the life cycle in addition to their direct effect previously shown by the initial specification. There is a clear increase in the magnitude of all the estimated coefficients associated to early-life conditions in the mediating specification compared with the results of model 1c. The effect of early-life conditions is magnified when mediating effects with other health determinants, such as education and lifestyles, are disentangled. The mediating specification 2 also emphasised that education level has both a direct and an indirect effect on health as the estimated coefficients associated to the education variable is larger. The mediating specifications allow us to evaluate the magnitude of the direct and indirect effects of early-life conditions on adult health, working through education and lifestyles.

Table III presents the magnitude of direct and indirect effects of each of the variables within the vector of early-life conditions. The absence of male head at the time of birth has the highest direct effect on adult health. The experience of financial hardships during childhood is also an important early-life condition influencing adult health directly. The absence of obesity at 16 years old has both relevant direct and indirect effects on adult health. As an indirect effect, it essentially works through lifestyles. The indirect effect of both parents' smoking also appears to work mainly through lifestyles but not significantly. Finally, individual educational attainment influences adult health both directly and indirectly through lifestyles.

4.4. Decomposition of health inequality

Table IV presents the results of the decomposition of the variance of the predicted latent health within the longitudinal panel data analysis for the alternative specifications (model 1c, specification 1 and specification 2).

The decomposition in the baseline specification shows that the most important contribution to health inequalities comes from the state dependence of health and the initial health, which would explain 35% of the variance in the predicted health. Lifestyles are directly explaining 32% of health inequalities, which confirm that they are important determinants of health inequalities. Early-life conditions explain about 13% of health inequalities. If we add their indirect contribution to health inequalities, as done in the mediating specification, the relative contribution of early-life conditions increases and would represent 19.2% of health inequalities. This increase underlines the existence of mediating effects of early-life conditions with education and lifestyles. On the contrary, the contribution of lifestyles on health inequalities reduces and would represent 27% of inequalities. Lifestyles are thus influenced by early-life conditions and, to a lesser extent, by educational level as their contribution to health inequalities slightly decreases down to 25% in the mediating specification 2. Comparison between the

DOI: 10.1002/hec

Table II. Random effect Probit models coefficients of the mediating specifications (with bootstrapped standard errors)

	Mediating specification 1			Mediating specification 2		
Variables	Coef.		SE	Coef.		SE
Gender (male)	0.086	**	0.044	0.086	**	0.045
Early-life conditions						
Fathers' social class (Ref.: I and II - Professional and managerial/technical)						
III – Skilled	−0.102		0.066	−0.102		0.066
IV and V – Partly skilled and unskilled	−0.279	***	0.077	−0.279	***	0.077
No male head	−0.465	***	0.115	−0.465	***	0.115
Financial hardship (Ref.: None)						
Yes	−0.348	***	0.072	−0.348	***	0.072
Non response	0.056		0.158	0.056		0.158
Father's education (Ref.: beyond the min age)						
Before or at the min age	−0.091		0.060	−0.091		0.060
Mother's education (Ref.: beyond the min age)						
Before or at the min age	−0.197	***	0.062	−0.197	***	0.062
Parental illness (Ref.: None)						
Father's illness	−0.191	**	0.080	−0.191	**	0.080
Mother's illness	−0.142		0.095	−0.142		0.095
Parental smoking (Ref.: None)						
Father's smoking	0.022		0.049	0.022		0.049
Non response	−0.026		0.095	−0.026		0.095
Mother's smoking	−0.124	***	0.045	−0.124	***	0.045
Non response	−0.080		0.132	−0.080		0.132
Chronic condition at 16 (Ref.: None)						
Yes	−0.059		0.065	−0.059		0.065
Non response	0.149		0.136	0.149		0.136
Low birth weight	−0.093		0.093	−0.093		0.093
Obesity at 16 (Ref.: Yes)						
No	0.183		0.189	0.183		0.189
Non response	−0.217	*	0.123	−0.217	*	0.123
Educational qualification						
Strictly lower than O-level	−0.169	***	0.054	−0.326	***	0.056
Lifestyles						
Lagged lifestyles						
Exercising once in the last 4 weeks	−0.045		0.046	−0.045		0.046
Non smoking	0.070		0.074	0.070		0.074
Drinking prudently	0.033		0.071	0.033		0.071
Non obese	−0.052		0.087	−0.052		0.087
Consumption of fruit/veg. once per week	0.072		0.096	0.072		0.096
Mean lifestyles						
Exercising once in the last 4 weeks	0.524	***	0.093	0.524	***	0.093
Non smoking	0.205	**	0.090	0.205	**	0.090
Drinking prudently	0.211	*	0.119	0.211	*	0.119
Non obese	0.771	***	0.129	0.771	***	0.129
Consumption of fruit/veg. once per week	0.318	***	0.123	0.318	***	0.123
Lagged health status	0.306	***	0.076	0.306	***	0.076
Initial conditions (health status at 23)	0.997	***	0.101	0.997	***	0.101
Time dummies (Ref.: t = 3)						
t = 1	0.579	***	0.047	0.579	***	0.047
t = 2	0.338	***	0.038	0.338	***	0.038
$V\left(\hat{H}^{*}\right)$	0.367			0.367		
σ_{ω}	0.636			0.636		
$\rho^{\#}$	0.389			0.389		
R^2 (McKelvey and Zavoina)	0.183			0.183		

[#]Share of individual unexplained heterogeneity measured by the share of σ_{ω} in the total unexplained variance $(\sigma_{\omega}+1)$.
Significance levels: ***1%, **5%, *10%.

DOI: 10.1002/hec

Table III. Direct and indirect effects of early-life conditions and education on adult health

Variables	Direct effect a_1			Indirect effect via education $b_1\varphi_1$			Indirect effect via lifestyles $c_1\varphi_1^a + c_2\varphi_1^b$			Total effect $a_1+\psi_1$		
	Coef.		**SE**	**Coef.**		**SE**	**Coef.**		**SE**	**Coef.**		**SE**
Early-life conditions												
Fathers' social class												
III – Skilled	−0.064		0.066	−0.005		0.013	−0.033		0.054	−0.102		0.066
IV and V – Partly skilled and unskilled	−0.206	***	0.076	−0.022		0.014	−0.052		0.055	−0.279	***	0.077
No male head	−0.383	***	0.109	−0.019		0.014	−0.063		0.055	−0.465	***	0.115
Financial hardship												
Yes	−0.253	***	0.070	−0.036	**	0.017	−0.059		0.055	−0.348	***	0.072
Non response	0.102		0.149	−0.013		0.013	−0.033		0.054	0.056		0.158
Father's education												
Before or at the min age	−0.046		0.061	−0.012		0.013	−0.034		0.054	−0.091		0.060
Mother's education												
Before or at the min age	−0.143	**	0.061	−0.015		0.013	−0.038		0.054	−0.197	***	0.062
Parental illness												
Father's illness	−0.165	**	0.078	−0.001		0.013	−0.025		0.054	−0.191	**	0.080
Mother's illness	−0.120		0.092	−0.001		0.012	−0.021		0.054	−0.142		0.095
Parental smoking												
Father's smoking	0.076	*	0.048	−0.005		0.013	−0.049		0.054	0.022		0.049
Non response	−0.009		0.093	0.001		0.013	−0.018		0.054	−0.026		0.095
Mother's smoking	−0.077	*	0.044	−0.013		0.013	−0.034		0.054	−0.124	***	0.045
Non response	−0.064		0.126	−0.010		0.013	−0.006		0.054	−0.080		0.132
Chronic condition at 16												
Yes	−0.016		0.065	−0.011		0.013	−0.032		0.054	−0.059		0.065
Non response	0.113		0.131	−0.010		0.013	0.045		0.055	0.149		0.136
Low birth weight	−0.074		0.089	−0.016		0.013	−0.003		0.054	−0.093		0.093
Obesity at 16												
No	−0.316	*	0.194	0.018		0.014	0.481	***	0.081	0.183		0.189
Non response	−0.153		0.118	−0.002		0.013	−0.061		0.055	−0.217	*	0.123
Educational qualification												
Strictly lower than O-level	−0.169	***	0.054				−0.158	***	0.056	−0.326	***	0.056

Significance levels: ***1%, **5%, *10%.

decompositions of the general specification and the mediating specifications suggest that the correlation between early-life conditions, and respectively lifestyles and education, is important. When we purge the contribution of lifestyles to health inequalities from their mediating effect with early-life conditions and education, we reduce their contribution to health inequalities and emphasise the importance of early-life conditions for health inequalities over the life cycle.

We thoroughly studied the share of early-life conditions in health inequalities and explore the relative contribution of social background, parents' health and lifestyles, and initial health to this vector. The decomposition in the general specification model suggests that social background variables are the leading contributing factor, representing about 79.4% of the share of early-life conditions in health inequalities.[8] Parents' health and lifestyles represent about 19.3%.

[8]Social background represents 10.15% of overall health inequality within the share of 12.79% represented by early-life conditions, so it represents 79.4% of the health inequality explained by early-life conditions.

DOI: 10.1002/hec

Table IV. Decomposition of health inequality (with bootstrapped standard errors and 95% confidence intervals using bootstrap percentiles method)

	Health inequality over the full period		
Variables	Share (%)	SE	(95% Confidence interval)
Baseline specification			
Demographics	16.55	2.21	(12.30–21.24)
Early-life conditions	12.79	2.46	(9.40–19.54)
Social background	*10.15*	*2.23*	*(6.24–14.96)*
Parents' health and lifestyles	*2.40*	*1.23*	*(0.96–5.73)*
Initial health	*0.24*	*0.68*	*(−0.39–2.08)*
Lifestyles	31.97	3.22	(25.83–37.32)
Education	4.03	1.56	(1.21–7.47)
Health state-dependence	34.66	3.35	(27.25–39.87)
Mediating specification 1			
Demographics	16.67	2.17	(12.51–21.25)
Early-life conditions	19.17	2.71	(15.24–25.91)
Social background	*14.11*	*2.46*	*(9.68–19.31)*
Parents' health and lifestyles	*3.68*	*1.40*	*(1.88–7.28)*
Initial health	*1.37*	*0.83*	*(0.39–3.46)*
Lifestyles	27.13	2.95	(21.85–32.51)
Education	2.37	1.07	(0.60–4.84)
Health state-dependence	34.66	3.35	(27.25–39.87)
Mediating specification 2			
Demographics	16.67	2.17	(12.51–21.25)
Early-life conditions	19.17	2.71	(15.24–25.91)
Social background	*14.11*	*2.46*	*(9.68–19.31)*
Parents' health and lifestyles	*3.68*	*1.40*	*(1.88–7.28)*
Initial health	*1.37*	*0.83*	*(0.39–3.46)*
Lifestyles	24.92	2.80	(20.04–30.43)
Education	4.58	1.41	(2.24–7.24)
Health state-dependence	34.66	3.35	(27.25–39.87)

5. CONCLUSION

In this paper, we developed a model to evaluate the contribution of several essential determinants of health to health inequalities using a representative cohort of individuals born in 1958 and a unique follow-up of health status, lifestyles as well as a good description of early-life conditions. Our results showed the indirect effects of early-life condition variables on health over the life cycle in addition to their direct effects. Early-life conditions have a larger contribution to health inequality when their indirect role on education achievement as an adult and lifestyles are taken into account, representing almost 20% of overall explained health inequality. This latter result underlines the relevance of mediating effects between the determinants of health and outperforms previous works excluding early-life conditions as a relevant determinant of health inequalities along with socioeconomic factors and lifestyles. Among early-life conditions, social background seems to be the most important determinant of overall health inequalities. Lifestyles show an impressive contribution to health inequalities, namely, 32% in the general specification but appear to be largely determined by both early-life conditions and education as their contribution reduced down to 25%. This confirms previous results, which have underlined social determinism along with the existence of an accumulation of risk on causal pathways in the life course (Kuh *et al*., 2003).

The study exhibits among the early-life conditions those which influence adult health directly, indirectly, and both directly and indirectly. The absence of father at the time of birth and experience of financial hardships represent the lead factors for direct effect on health. The absence of obesity at 16 influences health both directly and indirectly working through lifestyles. Indirect effects increase the relative contribution of

DOI: 10.1002/hec

early-life conditions to health inequalities and inversely decrease the contribution of lifestyles in the mediating model.

Finally, the dynamic panel analysis permits controlling a large part of individual unexplained heterogeneity as well as the important effect of health state dependence over time. Our study has some limitations. The inequality measure is based on the explained part of the variance that is allowed by the model specifications. According to the pseudo-R^2; that is built using the variance of the latent variable, we would be able to explain about 18%. Therefore, the unexplained health inequality remains very large. The panel data perspective also presents several limits. The first problem is the presence of attrition because of mortality in the cohort that we have ignored in the analysis. This leads us to an underestimation of the effect of early-life conditions, adult socioeconomic factors and lifestyles on health inequality as we worked on a selected sample of British people still alive at 46 years old. We did investigate mortality in our data, and we found that mortality rate appears to be more important before age 23 than between age 23 and age 46. Finally, the NCDS cohort has a singular structure as the different waves are not equidistant in time. In particular, there is a 4-year interval between the two last sweeps, whereas there were about 10 years between the past sweeps. We tried to catch this effect by introducing a year dummy into the models. Therefore, the estimated coefficients in the models can be interpreted as a mean of the effects of lifestyles, education and early-life conditions over time.

The study of the social determinants of health inequality together with lifestyles and the evaluation of their respective contribution to the magnitude of health inequality is particularly relevant for policy-makers. The legitimacy of policies to tackle health inequalities is related to the relative contribution of each broad factor. Early-life conditions, which are factors that cannot be chosen by the individual, appear to be relevant factors of health disparities. They are also considered as the most illegitimate sources of health inequalities by social justice philosophers such as Roemer (1998) and Dworkin (1981). Therefore, our results appear to orientate policy interventions towards the compensation of individual for unequal opportunities in health.

APPENDIX

SUPPLEMENTARY TABLES

Table A.I The original NCDS sample and the study sample

Year	**1958**	**1965**	**1969**	**1974**	**1981**	**1991**	**1999/2000**	**2004**
Cohort member age	Birth	7	11	16	23	33	42	46
Cross-sectional original sample	17,416	15,051	14,757	13,917	12,044	10,986	10,979	9,175
		Early-life conditions			***t = 0***	***t = 1***	***t = 2***	***t = 3***
Unbalanced selected sample					7,874	6,956	6,999	5,990
Balanced selected sample					4,480			

Table A.II Descriptive statistics in the balanced sample at t = 0

Variables	**N = 4,480**	**Proportion**
Gender		
Male	2,065	46.09%
Female	2,415	53.91%
Early-life conditions		

(*Continues*)

DOI: 10.1002/hec

Table A. II (Continued)

Fathers' social class		
I/II – Professional and managerial/technical	854	19.06%
III – Skilled	2,651	59.17%
IV/V – Partly skilled and unskilled	827	18.46%
No male head	148	3.30%
Financial hardship		
Yes	322	7.19%
No	4,055	90.51%
Non response	103	2.30%
Father's education		
Minimum schooling age and below	3,460	77.23%
Beyond the min age	1,020	22.77%
Mother's education		
Minimum schooling age and below	3,464	77.32%
Beyond the min age	1,016	22.68%
Parental illness		
Father's illness	314	7.01%
Mother's illness	218	4.87%
Parental smoking		
Father's smoking	2,432	54.29%
Non response	292	6.52%
Mother's smoking	1,962	43.79%
Non response	148	3.30%
Chronic condition at 16		
Yes	523	11.67%
No	3,475	77.57%
Non response	482	10.76%
Low birth weight	214	4.78%
Obesity at 16		
Yes	59	1.32%
No	3,818	85.06%
Non response	610	13.62%
Educational qualification		
O-level and higher	3,607	80.51%
Strictly lower than O-level	873	19.49%

Table A.III Distribution of health status in the balanced sample

	Age 23 *t = 0*	**Age 33** *t = 1*	**Age 42** *t = 2*	**Age 46** *t = 3*
Excellent	45.85%	35.51%	31.54%	32.08%
Good	46.88%	53.21%	53.19%	46.21%
Good health	**92.72%**	**88.73%**	**84.73%**	**78.28%**
Fair	6.70%	10.09%	12.77%	14.98%
Poor	0.58%	1.18%	2.50%	5.07%
Very poor				1.67%
Poor health	**7.28%**	**11.27%**	**15.27%**	**21.72%**

Table A.IV Descriptive statistics of lifestyles variables in the balanced sample

		Lifestyles	
	Age 23 *t = 0*	Age 33 *t = 1*	Age 42 *t = 2*
Exercising once in the last 4 weeks	49.51%	79.93%	75.56%
Non smoking	64.08%	71.52%	74.24%
Drinking prudently	87.61%	92.57%	84.98%
Non obese	97.37%	89.87%	85.65%
Consumption of fruit/veg. once per week	91.79%	91.79%	90.49%

DOI: 10.1002/hec

Table A.V Estimated coefficients (SD) of auxiliary equations (OLS model)

	Education	Exercising		Non smoking		Drinking prudently		Non obese		Eating fruit and vegetable	
Variables	**Before O-level**	**(a)**	**(b)**	**(a)**	**(b)**	**(a)**	**(b)**	**(a)**	**(b)**	**(a)**	**(b)**
Gender (male)	−0.057***	0.095***	0.088***	0.014	0.004	−0.088***	−0.088***	0.011*	0.009	−0.076***	−0.079***
	(0.011)	(0.009)	(0.009)	(0.012)	(0.012)	(0.007)	(0.007)	(0.006)	(0.006)	(0.007)	(0.007)
Early-life conditions											
Fathers' social class (Ref.: I and II - Professional and managerial/technical)											
III – Skilled	0.032*	−0.012	−0.009	−0.017	−0.011	0.011	0.011	−0.019**	−0.017**	−0.029***	−0.027***
	(0.017)	(0.013)	(0.013)	(0.016)	(0.016)	(0.010)	(0.010)	(0.008)	(0.008)	(0.009)	(0.009)
IV and V –Partly skilled and unskilled	0.128***	−0.052***	−0.036**	−0.048**	−0.011	0.025**	0.025**	−0.012	−0.007	−0.030**	−0.022**
	(0.021)	(0.016)	(0.016)	(0.021)	(0.016)	(0.012)	(0.012)	(0.011)	(0.011)	(0.012)	(0.012)
No male head	0.110***	−0.040	−0.027	−0.147***	−0.128***	0.011	0.011	0.001	0.005	−0.017	−0.010
	(0.034)	(0.026)	(0.026)	(0.038)	(0.038)	(0.020)	(0.020)	(0.018)	(0.018)	(0.020)	(0.020)
Financial hardship (Ref.: None)											
Yes	0.212***	−0.044**	−0.018	−0.087***	−0.050*	0.008	0.008	−0.013	−0.005	−0.018	−0.005
	(0.023)	(0.019)	(0.018)	(0.026)	(0.026)	(0.012)	(0.012)	(0.014)	(0.014)	(0.015)	(0.015)
Non response	0.077**	−0.027	−0.018	−0.021	−0.008	−0.022	−0.022	−0.022	−0.019	0.017	0.022
	(0.038)	(0.028)	(0.028)	(0.040)	(0.038)	(0.027)	(0.027)	(0.022)	(0.022)	(0.020)	(0.020)
Father's education (Ref.: beyond the min age)											
Before or at the min age	0.069***	−0.039***	−0.030**	−0.009	0.003	0.009	0.009	−0.014*	−0.011	−0.013	−0.008
	(0.016)	(0.012)	(0.012)	(0.016)	(0.016)	(0.010)	(0.010)	(0.008)	(0.008)	(0.009)	(0.009)
Mother's education (Ref.: beyond the min age)											
Before or at the min age	0.089***	−0.039***	−0.028**	−0.013	0.003	0.010	0.010	−0.017**	−0.013*	−0.018**	−0.012
	(0.015)	(0.011)	(0.011)	(0.015)	(0.015)	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)
Parental illness (Ref.: None)											
Father's illness	0.006	−0.001	0.000	−0.029	−0.028	0.012	0.012	−0.019	−0.019	−0.014	−0.014
	(0.022)	(0.018)	(0.018)	(0.025)	(0.024)	(0.012)	(0.012)	(0.014)	(0.014)	(0.015)	(0.015)
Mother's illness	0.003	−0.018	−0.018	−0.041	−0.040	−0.005	−0.005	0.004	0.005	−0.008	−0.008
	(0.026)	(0.021)	(0.020)	(0.029)	(0.029)	(0.016)	(0.016)	(0.014)	(0.014)	(0.017)	(0.018)
Parental smoking (Ref.: None)											
Father's smoking	0.031**	−0.017*	−0.013	−0.081***	−0.075***	−0.021***	−0.021***	−0.008	−0.007	−0.019**	−0.017**
	(0.012)	(0.009)	(0.009)	(0.013)	(0.013)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Non response	−0.007	0.015	0.014	−0.078***	−0.079***	−0.030*	−0.030*	0.014	0.014	−0.017	−0.018
	(0.025)	(0.019)	(0.018)	(0.027)	(0.027)	(0.016)	(0.016)	(0.013)	(0.013)	(0.015)	(0.015)
Mother's smoking	0.079***	−0.006	0.004	−0.044***	−0.030**	−0.008	−0.008	−0.020***	−0.017**	−0.007	−0.002
	(0.012)	(0.009)	(0.009)	(0.013)	(0.013)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Non response	0.061*	−0.025	−0.018	−0.007	0.003	0.021	0.021	0.008	0.011	−0.007	−0.004
	(0.032)	(0.025)	(0.024)	(0.036)	(0.034)	(0.017)	(0.017)	(0.017)	(0.017)	(0.022)	(0.022)
Chronic condition at 16 (Ref.: None)											
Yes	0.067***	−0.061***	−0.053	−0.004	0.008	0.018*	0.018*	−0.010	−0.007	0.005	0.009
	(0.018)	(0.014)	(0.014)	(0.019)	(0.019)	(0.010)	(0.010)	(0.011)	(0.011)	(0.011)	(0.011)

(*Continues*)

DOI: 10.1002/hec

Table A.V (Continued)

Variables	Education	Exercising		Non smoking		Drinking prudently		Non obese		Eating fruit and vegetable	
	Before O-level	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
Non response	0.059	0.047	0.054	0.030	0.020	0.022	0.022	−0.009	−0.006	0.039	0.043
	(0.037)	(0.030)	(0.029)	(0.043)	(0.042)	(0.022)	(0.022)	(0.019)	(0.019)	(0.026)	(0.027)
Low birth weight	0.094***	−0.021	−0.010	−0.006	0.011	0.030**	0.030**	0.009	0.013	−0.014	−0.008
	(0.026)	(0.021)	(0.021)	(0.029)	(0.029)	(0.013)	(0.013)	(0.015)	(0.015)	(0.017)	(0.017)
Obesity at 16 (Ref.: Yes)											
No	−0.104**	0.001	−0.012	0.083	0.065	−0.027	−0.027	0.640***	0.636***	0.011	0.005
	(0.049)	(0.035)	(0.036)	(0.053)	(0.055)	(0.027)	(0.027)	(0.043)	(0.043)	(0.033)	(0.034)
Non response	0.014	−0.051*	−0.049	−0.071*	−0.071*	−0.016	−0.016	0.007	0.007	−0.045*	−0.044*
	(0.034)	(0.027)	(0.027)	(0.039)	(0.038)	(0.021)	(0.021)	(0.017)	(0.017)	(0.024)	(0.025)
Educational qualification											
Strictly lower			−0.123***		−0.173***		0.000		−0.038***		−0.060***
than O-level			(0.012)		(0.017)		(0.008)		(0.010)		(0.011)
Time dummies (Ref.: t = 3)											
t = 1		−0.260***	−0.260***	−0.102***	−0.102***	0.026***	0.026***	0.117***	0.117***	0.013***	0.013***
		(0.009)	(0.009)	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)
t = 2		0.044***	0.044***	−0.027***	−0.027***	0.076***	0.076***	0.042***	0.042***	0.013***	0.013***
		(0.008)	(0.008)	(0.006)	(0.005)	(0.006)	(0.006)	(0.004)	(0.004)	(0.005)	(0.005)

Significance levels: *** 1%, **5%, *10%.
(a) Regression including early-life conditions;
(b) regression including early-life conditions and education.

DOI: 10.1002/hec

ACKNOWLEDGEMENTS

We gratefully acknowledge financial support from the Risk Foundation (Health, Risk and Insurance Chair, Allianz). Damien Bricard also benefited from a PhD studentship from the *Régime Social des Indépendants (RSI)*. Part of this work was carried out while Damien Bricard and Florence Jusot were visiting the Academic Unit of Health Economics at the University of Leeds in summer 2010 and while Sandy Tubeuf was a visiting professor at the Université Paris-Dauphine in autumn 2010. We thank colleagues from the Academic Unit of Health Economics and the Leeds Institute of Health Sciences for their helpful comments during internal seminars. We are grateful to two anonymous referees, Anne-Laure Samson, Gang Chen, Silvia Balia, Brenda Gannon, Christine Le Clainche, and participants to the JESF 2010, the LEDA-LEGOS seminar, the Second Australasian Workshop on Econometrics and Health Economics, the CORE workshop on Equity in Health, the JMA 2011, and the 8th iHEA Congress for useful comments and suggestions. The National Child Development Study data were supplied by the UK Data Archive (usage 54452); the authors alone are responsible for the analysis and interpretation in this paper.

CONFLICT OF INTERESTS

The authors declare no conflict of interests.

REFERENCES

Anda RF, Whitfield CL, Felitti VJ, Chapman D, Edwards VJ, Dube, SR, Williamson DF. 2002. "Adverse childhood experiences, alcoholic parents, and later risk of alcoholism and depression", *Psychiatric Services* **53**(8): 1001–1009.

Arulampalam W, Stewart MB. 2009. Simplified implementation of the Heckman estimator or the dynamic Probit model and a comparison with alternative estimators. *Oxford Bulletin of Economics and Statistics* **71**(5): 659–681.

Balia S, Jones A. 2008. Mortality lifestyles and socio-economic status. *Journal of Health Economics* **27**: 1–26.

Bernt Karlson K, Holm A. 2011. Decomposing primary and secondary effects: a new decomposition method. *Research in Social Stratification and Mobility* **29**: 221–237.

Bernt Karlson K, Holm A, Breen R. 2012. Comparing Regression Coefficients Between Same-Sample Nested Models using Logit and Probit: A New Method. *Sociological Methodology* **42** [Epub ahead of print].

Case A, Katz L. 1991. The company you keep: the effects of family and neighborhood on disadvantaged youths. *NBER Working Paper*.

Case A, Fertig A, Paxson C. 2005. The lasting impact of childhood health and circumstance. *Journal of Health Economics* **24**: 365–389.

Contoyannis P, Jones A. 2004. Socio-economic status, health and lifestyle. *Journal of Health Economics* **23**: 965–995.

Currie J, Stabile M. 2003. Socioeconomic status and child health: why is the relationship stronger for older children. *American Economic Review* **93**(1813): 1823.

Cutler DM, Deaton A, Lleras-Muney A. 2006. The determinants of mortality. *Journal of Economic Perspectives* **20**(3): 97–120.

de Walque D. 2007. Does education affect smoking behaviors? Evidence using the Vietnam draft as an instrument for college education. *Journal of Health Economics* **26**(5): 877–895.

Dworkin R. 1981. What is equality ? Part I: Equality of Welfare. *Philosophy and Public Affairs* **10**: 185–246.

Etilé F, Jones A. 2011. Schooling and smoking among the Baby Boomers: an evaluation of the impact of educational expansion in France. *Journal of Health Economics* **30**(4): 811–831.

Francesconi M, Jenkins SP, Siedler T. 2010. The effect of lone motherhood on the smoking behavior of young adults. *Health Economics* **19**(11): 1377–1384.

Francesconi M & Jenkins SP & Stephen P, Siedler T. 2010. "The effect of lone motherhood on the smoking behavior of young adults", *Health Economics*, John Wiley & Sons, Ltd., **19**(11): 1377–1384.

Frisvold D, Golberstein E. 2011. School quality and the education-health relationship: evidence from Blacks in segregated schools. *Journal of Health Economics* **30**(6): 1232–1245.

Göhlmann S, Schmidt CM, Tauchmann H. 2010. Smoking initiation in Germany: the role of intergenerational transmission. *Health Economics* **19**(2): 227–242.
Grimard F, Parent D. 2007. Education and smoking: were Vietnam war draft avoiders also more likely to avoid smoking? *Journal of Health Economics* **26**(5): 896–926.
Grossman M. 2006. Education and nonmarket outcomes. In *Handbook of the Economics of Education - vol 2*, Hanushek EA, Welch F (eds.), Amsterdam: North-Holland, Elsevier edn: 577–633.
Hakkinen U, Jarvelin M-R, Rosenqvist G, Laitinen J. 2006. Health, schooling and lifestyle among young adults in Finland. *Health Economics* **15**(11): 1201–1216.
Heckman JJ. 1981. The incidental parameter problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. In *Structural Analysis of Discrete Data with Econometric Applications*, Manski CF, McFadden DL (eds), the MIT Press edn: Cambridge; 179–195.
Idler EL, Benyamini Y. 1997. Self-rated health and mortality: a review of twenty-seven community studies. *Journal of Health and Social Behaviour* **38**(March): 21–37.
Jürges H, Reinhold S, Salm M. 2011. Does schooling affect health behavior? Evidence from the educational expansion in Western Germany. *Economics of Education Review* **30**(5): 862–872. Special Issue on Education and Health.
Jusot F, Tubeuf S, Trannoy A. 2010. Effort or circumstances: does the correlation matter for inequality of opportunity in health. *Working Paper IRDES*, vol. DT33.
Kemptner D, Jürges H, Reinhold S. 2011. Changes in compulsory schooling and the causal effect of education on health: Evidence from Germany. *Journal of Health Economics* **30**(2): 340–354.
Kenkel DS. 1991. Health behavior, health knowledge, and schooling. *Journal of Political Economy* **99**(2): 287–305.
Khang Y-H, Lynch J, Yang S, Harper S, Yun S-C, Jung-Choi K, Kim HR. 2009. The contribution of material, psychosocial, and behavioral factors in explaining educational and occupational mortality inequalities in a nationally representative sample of South Koreans: relative and absolute perspectives. *Social Science and Medicine* **68**(5): 858–866.
Kjellsson G, Gerdtham UG, Lyttkens CH. 2011. Breaking bad habits by education - smoking dynamics among Swedish women. *Health Economics* **20**(7): 876–881.
Kohler U, Bernt Karlson K, Holm A. 2011. Comparing coefficients of nested nonlinear probability models. *The Stata Journal* **11**(3): 420–438.
Koivusilta LK, Rimpelä A, Rimpelä MK. 1999. Health-related lifestyle in adolescence-origin of social class differences in health? *Health Education Research* **14**(3): 339–355.
Koivusilta L, Rimpelä A, Vikat A. 2003. Health behaviours and health in adolescence as predictors of educational level in adulthood: a follow-up study from Finland. *Social Science & Medicine* **57**(4): 577–593.
Kuh D, Ben-Shlomo Y, Lynch J, Hallqvist J, Power C. 2003. Life course epidemiology. *Journal of Epidemiology and Community Health* **57**(10): 778–783.
Laaksonen M, Talala K, Martelin T, Rahkonen O, Roos E, Helakorpi S, Laatikainen T, Prattala R. 2008. Health behaviours as explanations for education level differences in cardiovascular and all-cause mortality: a follow-up of 60,000 men and women over 23 years. *European Journal of Public Health* **18**(1): 38–43.
Lahti-Koski M, Gill T. 2004. Defining Childhood Obesity. In *Obesity in Childhood and Adolescence*, Kiess W, Marcus C, Wabitsch M (eds.), Pediatr Adolesc Med.: Basel, Karger, 1–19.
Laitinen J, Power C, Järvelin M-R. 2001. Family social class, maternal body mass index, childhood body mass index, and age at menarche as predictors of adult obesity. *American Journal of Clinical Nutrition* **74**(3): 287–294.
Lantz P, Golberstein E, House JS, Morenoff J. 2010. Socioeconomic and behavioral risk factors for mortality in a national 19-year prospective study of US adults. *Social Science & Medicine* **70**(10): 1558–1566.
Lindeboom M, Llena-Nozal A, van der Klauuw B. 2009. Parental education and child health: evidence from a schooling reform. *Journal of Health Economics* **28**(1): 109–131.
Marmot M, Shipley MJ, Brunner EG, Hemingway H. 2001. Relative contributions of early-life and adult socioeconomic factors to adult morbidity in the Whitehall II study. *Journal of Epidemiology and Community Health* **55**(5): 301–307.
McKelvey RD, Zavoina W. 1975. A statistical model for the analysis of ordinal level dependent variable. *Journal of Mathematical Sociology* **4**: 103–120.
Menvielle G, Boshuizen H, Kunst AE, Dalton S, Vineis P, Bergmann M, Silke H, Ferrari P, Raaschou-Nielsen O, Tjonneland A, Kaaks R, *et al*. 2009. The role of smoking and diet in explaining educational inequalities in lung cancer incidence. *Journal of the National Cancer Institute* **101**(5): 321–330.
Miranda A. Dynamic probit models for panel data: A comparison of three methods of estimation. 2007. United Kingdom Stata Users' Group Meetings 2007 11. Ref Type: Slide
Moser K, Li L, Power C. 2003. Social inequalities in low birth weight in England and Wales: trends and implications for future population health. *Journal of Epidemiology and Community Health* **57**: 687–691.
Mundlak Y. 1978. On the pooling of time series and cross-section data. *Econometrica* **46**(1): 69–85.

DOI: 10.1002/hec

Oreopoulos P. 2006. "Estimating average and local average treatment effects of education when compulsory schooling laws really matter", *American Economic Review* **96**(1): 152–175.
Orme C. 2001. Two-step inference in dynamic non-linear panel data models. *Mimeo* no. Revision of Economics Discussion Paper 9633: "The initial conditions problem and two-step estimation in discrete panel data models".
van Ours JC, Williams J. 2009. Why parents worry: initiation into cannabis use by youth and their educational attainment. *Journal of Health Economics* **28**(1): 132–142.
Park C, Kang C. 2008. Does education influence healthy lifestyles? *Journal of Health Economics* **27**(6): 1516–1531.
Power C, Graham H, Due P, Hallqvist J, Joung I, Kuh D, *et al*., 2005. The contribution of childhood and adult socioeconomic position to adult obesity and smoking behaviour: An international comparison. *International Journal of Epidemiology* **34**; 335–344.
Reinhold S, Jürges H. 2010. Secondary school fees and the causal effect of schooling on health behavior. *Health Economics* **19**(8): 994–1001.
Roemer J. 1998. *Equality of opportunity*. Harvard University Press edn: Cambridge.
Rosa-Dias P. 2009. Inequality of Opportunities in Health: Evidence from a UK cohort study. *Health Economics* **18**(9): 1057–1074.
Shorrocks A. 1982. Inequality Decomposition by Factor Components. *Econometrica* **50**(1): 193–211.
Silles MA. 2009. The causal effect of education on health: Evidence from the United Kingdom. *Economics of Education Review* **28**(1): 122–128.
Skalicka V, van Lenthe F, Bambra C, Krokstad S, Mackenbach JP. 2009. Material, psychological, behavioural and biomedical factors in the explanation of relative socio-economic inequalities in mortality: evidence from the HUNT study. *International Journal of Epidemiology* **38**(5): 1272–1284.
Staff J, Patrick M, Loken E, Maggs J. 2008. Teenage alcohol use and educational attainment. *Journal of Studies on Alcohol and Drugs* **69**(6): 848–858.
Strand BH, Tverdal A. 2004. Can cardiovascular risk factors and lifestyle explain the educational inequalities in mortality from ischaemic heart disease and from other heart diseases? 26 year follow up of 50 000 Norwegian men and women. *Journal of Epidemiology and Community Health* **58**(8): 705–709.
Stringhini S, Sabia S, Shipley MJ, Brunner EG, Nabi H, Kivimaki M, Singh-Manoux A. 2010. Association of socioeconomic position with health behaviors and mortality. *Journal of the American Medical Association* **303**(12): 1159–1166.
Trannoy A, Tubeuf S, Jusot F, Devaux M. 2010. Inequality of Opportunities in Health in France: A first pass. *Health Economics* **19**(8): 921–938.
van Doorslaer E, Koolman X. 2004. Explaining the differences in income-related health inequalities across European countries. *Health Economics* **13**(7): 609–628.
Van Oort F, van Lenthe F, Mackenbach JP. 2005. Material, psychosocial, and behavioural factors in the explanation of educational inequalities in mortality in the Netherlands. *Journal of Epidemiology and Community Health* **59**(3): 214–220.
von Hinke Kessler Scholder S. 2008. Maternal employment and overweight children: does timing matter? *Health Economics* **17**(8): 889–906.
Webbink D, Martin NG, Visscher PM. 2010. Does education reduce the probability of being overweight? *Journal of Health Economics* **29**(1): 29–38.
Wooldridge J. 2002. *Econometric Analysis of Cross Section and Panel Data*. MIT Press edn: Cambridge.
Working Party of the Royal College of Physicians UK. 2001. *Alcohol - Can the NHS afford it? Recommendations for a coherent alcohol strategy for hospitals*. NHS edn: London.

DOI: 10.1002/hec