HEDG Working Paper 08/18

Modelling and Measuring Inequality of Opportunity in Health: Evidence from a Cohort Study

Pedro Rosa Dias

May 2008

ISSN 1751-1976

http://www.york.ac.uk/res/herc/research/hedg/wp.htm

Modelling and Measuring Inequality of Opportunity in Health: Evidence from a Cohort Study

Pedro Rosa Dias[1]

May 2008

(Preliminary version: please do not quote)

Abstract

This paper uses data from the National Child Development Study to propose an empirical implementation of the concept of inequality of opportunity in health. Drawing on the distinction between *circumstance* and *effort* variables in John Roemer's work on equality of opportunity, *circumstances* are proxied by parental socio-economic status and childhood health; *effort* is proxied by health-related lifestyles and educational attainment. The paper is divided in three parts. First, a set of stochastic dominance tests is used to detect inequality of opportunity in the conditional distributions of self-assessed health in adulthood. Second, relying on a comprehensive set of circumstances, two alternative approaches are used to measure inequality of opportunity in health. Finally, in order to illuminate the triangular relationship between circumstances, effort and health, a structural model which relates self-assessed health in adulthood to lifestyles and educational attainment is considered. A recursive system of equations for self-assessed health, lifestyles and educational attainment is estimated by full information maximum likelihood to unveil the causal relationships at stake. The results indicate the existence of considerable and persistent inequality of opportunity in health. They also suggest that circumstances affect health in adulthood both directly and through effort factors such as educational attainment. This indicates that, while the influence of some unjust circumstances can only be tackled during childhood, the implementation of complementary educational policies may be of paramount importance to reduce health inequalities.

[1] Department of Economics and Related Studies & Centre for Health Economics, The University of York. Correspondence: Centre for Health Economics, Alcuin “A” Block, University of York, Heslington, York, YO10 5DD, United Kingdom. Tel: +44 1904321411, E-mail:spcrd500@york.ac.uk.

1. Introduction

Much of the attention traditionally given to equality of outcomes has shifted towards equality of opportunities. This change of emphasis is the consequence of the latest developments in political philosophy, inspired by the work of Rawls and Sen, systematised by Dworkin (1981), and subsequently modified by Arneson (1989) and Cohen (1989). In recent years, equality of opportunity prompted a series of applications in different fields of economic research[2] and attracted growing interest of policy makers, as becomes clear in the World Bank Development Report 2006. Within health economics, Rosa Dias and Jones (2007) argued that equality of opportunity is the implicit underlying concept of a broad range of inequality studies published over the last decade. Despite of this, the number of empirical applications of this concept to health is still scarce[3]; this paper aims primarily at narrowing this gap.

All conceptions of equal opportunity draw on some distinction between fair and unfair sources of inequality. Environmental factors such as genetic endowment and parental income are largely seen as illegitimate sources of health inequalities. On the contrary, the differences in health status that are due to lifestyles, are often seen as ethically justified by individual free choice. These contrasting sorts of factors have been studied autonomously by two well developed strands of research: the literature on the impact of childhood conditions on adult health and the research concerned with health and lifestyles. The interaction between these two is however much less explored. Furthermore, both strands were developed in relative isolation from the literature on health inequalities. Establishing a bridge between all these branches of research is the second purpose of this paper.

The paper is grounded on the framework proposed by Roemer (1998, 2002). This is then augmented with a set of testable condition defined in Lefranc et al. (2005, 2006) and embedded in the structural modelling approach of Fleurbaey and Schokkaert (2007). The data used are from the National Child Development Study (NCDS).

The results indicate that the data are consistent with the presence of inequality of opportunity in health. Two alternative approaches for its measurement are proposed and discussed. The paper also identifies the childhood circumstances that determine health in adulthood and uses a structural model to illuminate the channels by which these circumstances impact on health: direct channels and indirect pathways, influencing health through lifestyles. Two policy implications are inferred. First, there are causes of inequality of opportunity that can only be effectively tackled in childhood. An example of these is childhood obesity, which has been targeted by specific policy interventions in recent years. Second, some key policies to reduce inequality of opportunity in health may be found outside the health care system. The influence of education on lifestyles, which strongly affect health, suggests that complementary educational policies may be of paramount importance.

---

[2] For example Betts and Roemer (2001), Le Grand et al. (2002), Dardanoni et al. (2004), Lefranc et al. (2005) and Bourguignon et al. (2005).

[3] Devaux et al. (2008) is one of the very few recent empirical papers focused on inequality of opportunity in health.

2. Background

2.1 Equality of Opportunity: the Roemer model

The empirical analysis developed in this paper is explicitly grounded on the theoretical framework of the Roemer model (1998, 2002). It starts by sorting all factors influencing individual attainment between a category of *effort factors*, for which individuals should be held responsible and a category of *circumstance factors*, which, being beyond individual control, are the only source of illegitimate differences in outcomes. The outcome of interest is health as an adult (*H*); a health production function $H(C,E)$ is defined, where C denotes individual circumstances and E denotes effort[4].

The Roemer model does not specify which causal factors constitute circumstances and effort. In the case of inequality of opportunity in health, this dilemma is facilitated by the existence of medical and economic evidence on the main determinants of health in adulthood. There is a branch of economic literature devoted to the impact of childhood circumstances on health outcomes: Currie and Stabile (2004), Case et al. (2005) and Lindeboom et al. (2006) are recent examples. Using different datasets, these studies appraise conflicting theories about the channels by which childhood cconditions influence long term health and labour market outcomes. The most prominent among these theories are: the *fetal-origins hypothesis* (Barker (1995), Raveli et al (1998)) according to which parental socioeconomic characteristics influence the *in utero* conditions for fetal growth which, in turn, condition long term health; the *life course models* (Kuh and Wadsworth (1993)) which emphasise the impact of deprivation in childhood on adult health and longevity; the *pathways models* (Marmot et al. (2001)) which suggest that health in early life is important only because it will condition the socioeconomic position in early adulthood, which explains disease risk latter in life.

This paper follows this strand of research: it considers as circumstances the parental socioeconomic characteristics, spells of financial hardship during the cohort members' childhood and adolescence, proxies of congenital endowment such as the prevalence of chronic conditions in the family and birth weight, as well as incidence of acute conditions, chronic illnesses and obesity in childhood and early adolescence. All these factors affect the cohort members before the age of 16, reflecting conditions and choices that are largely beyond individual control.

There is also considerable work done on the relationship between health and lifestyles; recent examples include Mullahy and Portney (1990), Kenkel (1995), Contoyannis and Jones (2004) and Balia and Jones (2007). Lifestyles, such as cigarette smoking, alcohol consumption, diet and physical exercise are at least partially within individual control, hence they constitute the primary effort factors in this paper. Since educational attainment is also partially within individual responsibility and given that it is a plausible explanatory factor of health in adulthood, it is also taken here for an effort factor; this is however not an essential assumption.

[4] In rigorous terms, following the Roemer specification would imply defining a health production function $H(C,E,\Phi)$, where Φ stands for the action of public policy. Since no specific policy is considered in this paper, the health production function is simplified to $H(C,E)$.

The Roemer model defines social types consisting of the individuals who share exposure to the same circumstances. The set of observed individual circumstances allows the specification of these social types within the data. It is assumed that the society has a finite number of T types and that, within each type, there is a continuum of individuals. A fundamental aspect in this setting, is the fact that the distribution of effort within each type (F^t) is itself a characteristic of that type; since this is beyond from individual control, it constitutes a circumstance.

In order to make the degree of effort expended by individuals of different types comparable, Roemer proposes the definition of quantiles of the effort distribution (in this case, the number of cigarettes per day or number of units of alcohol consumed per week) within each type: two individuals are deemed to have exerted the same degree of effort if they sit at the same quantile (π) of their type's distribution of effort. When effort is observed, this definition is directly applicable. However, if effort is either unobservable or incommensurable, an additional assumption is required: by assuming that the outcome, health in this case, is monotonically increasing in effort, i. e. that healthy lifestyles are a positive contribution to the health stock, effort becomes the residual determinant of health once types are fixed; therefore, those who sit at the π^{th} quantile of the outcome distribution also sit at the π^{th} quantile of the distribution of effort within his type[5]. Both versions of the model are compatible with the empirical analysis done in this paper.

The definition of equality of opportunity used in this paper also follows from the Roemer model: equality of opportunity attains when outcomes are identical across types at fixed levels of effort. In the case of equal opportunity for health, this means that, on average, all those who adopt identical lifestyles should be entitled to experience a similar health status, irrespective of their circumstances. Such a situation corresponds to a full nullification of the effect of circumstances, keeping untouched the differences in outcome that are caused solely by effort[6].

2.2 Definitions and testable conditions

The definition of equality of opportunity given by Roemer (2002) is more appropriate for the situation in which a public policy is being evaluated rather than for inequality measurement from survey data. A set of alternative definitions was recently proposed by Lefranc et al. (2006) and Devaux et al (2008): these appeal to the concept of stochastic dominance and are coherent with the rationale of the previous section.

---

[5] It must be noted that this assumption comes at a cost: attributing the entire variation of outcome conditional on circumstances to effort, one is implicitly assuming that all relevant circumstances have been taken into account. If this fails to hold, measures of inequality of opportunity must be seen as lower bounds. To overcome this problem, Lefranc et al. (2006) propose that a third factor of *luck* be added to circumstances and effort in order to refine inequality measurement. In empirical terms, simply adding *luck* does not require adaptations: if *luck* it is orthogonal to circumstances, then it is innocuous; if it is correlated to them, then it must is treated as unobserved heterogeneity which should require the researcher's attention whether or not luck is explicitly considered.

[6] Roemer (1998; 2002) imposes an additional assumption in order to define more formally equality of opportunity: it is assumed that effort may be represented by a scalar, whence an indirect outcome function $v^t(\pi,\phi)$ may be defined for each type, depending on effort and on social policy. The model is then solved for the equal-opportunity policy ϕ^* for which $v^t(\pi,\phi)$ are equal across types at fixed levels of π. Since this paper does not aim at policy evaluation, this assumption is of little help here.

The use of stochastic dominance in the field of economic inequality is well established: as made clear by Sen and Foster (1997; pp 138 - 139), there are two main approaches to the construction of unanimity quasi-orderings of income distributions. One involves integrating the inverse of cumulative distribution functions (cdf) along the *population axis*, defining generalised Lorenz curves. The other, starts by integrating *cdfs* along the *income axis*, defining stochastic dominance relationships that are extensively used in financial economics.

A lottery stochastically dominates another if it yields a higher expected utility. Several orders of stochastic dominance may therefore be defined according to the restrictions one is willing to make on the individual utility function. First order stochastic dominance (FSD) holds for the whole class of increasing utility functions ($u'>0$); this corresponds to simply comparing *cdfs* of the earnings paid by alternative lotteries. Second order stochastic dominance (SSD) applies to utility functions which are increasing and concave in income, reflecting the notion of risk aversion ($u'>0$ and $u''<0$); SSD evaluates integrals of the *cdfs*. While FSD implies SSD, the converse is clearly not true.

These assumptions define broad classes of utility functions and are therefore applicable to the case of health. The exposure to different circumstances defines alternative lotteries; stochastic dominance allows the comparison of their health-related outcomes under standard assumptions on preferences. Also, contrasted with the benchmark of the Lorenz curve / concentration curve analysis of health inequalities, no additional assumptions are imposed by stochastic dominance, both at first and second orders.

Roemer's notion of inequality of opportunity applies to individuals who, having expended the same effort, achieve different outcomes due to different circumstances; inequalities due to effort are deemed acceptable. Denoting by *F(.)* the *cdf* of health, a literal translation of this would mean saying that there is inequality of opportunity whenever: $\forall (c;c')\,\forall e, F(.\,|\,c,e) \neq F(.\,|\,c',e)$.

This condition is however too stringent to be useful in empirical work. Lefranc et al. [7] (2006) consider that the data are consistent with the hypothesis of inequality of opportunity if the social advantage provided by different circumstances can be unequivocally ranked by SSD, i.e. if the distributions of health conditional on different circumstances can be ordered according to expected utility:

$$\forall c \neq c', \forall e, F(.\,|\,c,e) \succ_{SSD} F(.\,|\,c',e).$$

In this paper the main outcome of interest is self-assessed health, which is inherently ordinal. This fact dictates the need of redefining this condition in terms of FSD:

$$\forall c \neq c', \forall e, F(.\,|\,c,e) \succ_{FSD} F(.\,|\,c',e).$$

Since FSD implies SSD, this is a stronger condition, which necessarily satisfies the requirements set by Lefranc et al. (2006). It is also statistically testable and therefore it is used to assess the existence of inequality of opportunity in section 4.

[7] Lefranc et al. (2006) propose a wide range of stochastic dominance definitions of equality of opportunity which apply depending on whether circumstances and effort are observed. The choice of the definition above is motivated by the characteristics of the data and of the framework of the Roemer model: absent individual heterogeneity, circumstances are always assumed to be observed and effort is either assumed to be observed or positively correlated with the health outcomes.

2.3 Measures of inequality of opportunity

The stochastic dominance conditions are testable, but do not provide a measure of inequality of opportunity in health. For this purpose, this paper uses two alternative measures. The first is the Gini-opportunity index, first put forward by Lefranc et al. (2005). It quantifies the health inequality between different social types, defined by the researcher according to the exposure to particular circumstances. The second is a measure that avoids the subjective definition of types and which constitutes a variation of the *egalitarian-equivalence approach* proposed by Fleurbaey and Schokkaert (2007).

2.3.1 The Gini-opportunity index

The area underneath the generalised Lorenz curve (*A*) relates to the Gini coefficient according to: $A = \int GL(p)dp = \frac{1}{2}\mu(1-G)$. The double of *A*, i.e. the expression $\mu(1-G)$, is known as the *Sen evaluation function*[8], and constitutes the primary measure of social welfare when only the mean level of outcome and the Gini coefficient[9] are known.

In this context, Bensaid and Fleurbaey (2003) interpret the area underneath the generalised Lorenz curve as a cardinal measure of opportunity: for example, the area underneath the generalised Lorenz curve of one given type is a measure of that type's opportunity set. Following this line of thought, Lefranc et al. (2005) propose using a modified Gini coefficient to quantify the inequality between the different types' opportunity sets: ranking types (not individuals) according to their respective values of $A_j = \mu_j(1-G_j)$ and starting from the smallest one, the *Gini-Opportunity index* is defined as: $G-Opp = \frac{1}{\mu}\sum_{i}^{k}\sum_{i<j} p_i p_j \left[\mu_j\left(1-G_j\right) - \mu_i\left(1-G_i\right)\right]$.

This index, gives the weighted average of the differences between the types' opportunity sets in which the weights are the sample weights of the different types ($p_{i,j}$). It must be noticed that Gini-opportunity index increases in the number of types, therefore depending on the subjective definition of these by the researcher[10].

2.3.2 The *egalitarian-equivalence* approach

In some situations, the definition of social types has a clear intuitive appeal; in others, however, it may be hard to justify. In order to avoid this downside, one may treat each individual as a type: by assuming that the number of social types equals the number of

---

[8] There are several ways of interpreting the Sen evaluation function in terms of social welfare. These range from the original Sen's pairwise maximin criterion, to the grounds of relative deprivation and its relationship with a particular class of altruistic welfare functions. These connections are meticulously discussed in Lambert (2001, p. 122-26).

[9] It is not possible to summarise a social welfare function using solely the mean income and the Gini coefficient. The link between the Gini coefficient and social welfare has therefore to be made in non-individualistic terms.

[10] The Gini-opportunity index also satisfies all the fundamental properties required by the indices of relative inequality: within type anonymity; between-type Pigou-Dalton principle of transfers; normalisation (if cdfs are equal, the index is equal to zero); homogeneity of degree zero; invariance to a replication of the population. For details see Lefranc et al. (2005) and references therein.

individuals, the Gini-opportunity index equals, by construction, the conventional Gini coefficient.

Although the health concentration index has been the workhorse in much of the recent work on health inequalities, Fleurbaey and Schokkaert (2007) propose an approach to the measurement of unfair inequalities that uses primarily the Gini coefficient. This framework encompasses the measurement of inequality of opportunity and, under the assumption that the number of social types equals the number of individuals, it avoids the subjective definition of types.

Fleurbaey and Schokkaert (2007: pp 8) define unfair health inequalities on the grounds of a *compensation principle*: inequality is zero if two individuals with the same value for the effort variables are able to reach the same outcome. Under the assumption that the number of types equals the number of individuals, this corresponds to the definition of inequality of opportunity in the Roemer model. In the light of this principle, Fleurbaey and Schokkaert (2007) propose an *egalitarian equivalence method* for measuring health inequalities; this consists of applying the Gini coefficient to the vector $[h_i - h_i^*]$, where h_i denotes the observed health distribution and h_i^* stands for a hypothetical distribution of health outcomes in which the value of the circumstances is fixed to remove all the illegitimate sources of variation.

In this paper, a similar approach is used. The health outcome is indirectly standardised for circumstances by running $h_i = \alpha + \beta C_i + \varepsilon_i$ and computing $\hat{h}_i = \hat{\beta} C_i = h_i - \hat{\varepsilon}_i$. The pseudo-Gini coefficient[11] is then applied directly to $\hat{h}_i$, in order to measure the overall health inequality that is due to circumstances, hence the extent of inequality of opportunity.

This approach diverges from Fleurbaey and Schokkaert (2007) with respect to the indirect standardisation procedure: the first stage regression implemented in this paper omits all the effort variables; as pointed-out by Gravelle (2003), this might lead to biased estimates, for the partial correlations between circumstances are not taken into account. However, in the context of the Roemer model, these partial correlations should also be treated as circumstances for they embody the indirect effect of the unjust circumstances on health that is channelled through effort. This omission is therefore deliberate.

The value of this simple measure is directly comparable with that of the conventional health pseudo-Gini coefficient $G(h_i)$. The health pseudo-Gini coefficient has been used in the literature to measure inequality of outcomes. It implicitly treats as circumstances all the sources of variation in health and, therefore, the value of $G(h_i)$ constitutes an upper bound for inequality of opportunity. In turn, $G(\hat{h}_i)$ treats as circumstances only the sources of unfair inequality that are labelled as such by the researcher; it is therefore a lower bound for the extent of inequality of opportunity in health.

---

[11] The outcome of interest in this paper is self-assessed health, measured in a discrete ordinal scale. Because of this, individuals cannot be simply ranked by health: grouped data is therefore used and *pseudo-Lorenz curves* and *pseudo-Gini coefficients* defined. A rigorous treatment of these inequality measurement tools can be found in Van Doorslaer and Jones (2003, pp. 68-70).

2.4 Structural model: a normative interpretation of the Grossman model

Levels of effort are the consequence of utility maximisation subject to constraints: this issue is however left out by the Roemer model. Fleurbaey and Schokkaert (2007) propose the formulation of a behavioural model to explain the interaction between legitimate and illegitimate sources of inequality: this illuminates the channels by which circumstances affect health outcomes. Given the nature of the data used here, one may add that such a model also allows one to link the literature on childhood circumstances and the work done on health and lifestyles, which have evolved in relative isolation. The structural model put forward in this paper is a normative interpretation of Grossman's (1972) seminal model; the specification that follows draws heavily on recent variants of this model such as Adda and Lechene (2001), Contoyannis and Jones (2004) and Balia and Jones (2007).

Following the Grossman (1972) model, it is assumed that health is simultaneously a consumption good and a fundamental commodity[12] produced by inputs that are labelled circumstances or effort by the researcher. Effort factors are choice variables by definition and their marginal product is assumed to be known to the individuals. Each individual maximises lifetime utility, subject to income and time constraints, assuming that the time of death is uncertain: the probability of surviving from time *t* to time *t+1* is known and denoted π_t. The maximisation problem is:

$$\max \sum_{t=0}^{\infty} \beta^t \pi_t U\left(E_t, H_t, C_t, \mu_u\right) \qquad (1)$$

At each time period, utility depends on effort and circumstance factors (circumstances are exogenous by definition), on the health state variable and on factors unobserved by the researcher although arguably known to the individual (μ_u): genetic propensities, for example, are circumstances that may condition effort responses aimed at offsetting risks of illness, but which are hidden to the researcher. The instantaneous utility is discounted according to a subjective discount factor (β) and to the probability of survival until the next period (π_t).

For completeness, it is also assumed that the law of motion of the health stock is:

$$H_{t+1} = f(E_t, C_t, \mu_h) + \left(1-\delta\right) H_t \qquad (2)$$

As in the original Grossman model, the health production function is assumed to be increasing and concave in effort and the depreciation rate of the health stock (δ) is positive and smaller than unity.

Assume that individuals maximise (1) subject to an income constraint and to a time constraint:

[12] Two aspects deserve clarification. First, health constitutes a *fundamental commodity* in the sense that it is an argument of the (direct) utility function; no ethical judgment is attached to this assumption. Second, the literature encompasses more refined versions of Grossman model than the one presented here: Dardanoni and Wagstaff (1987) and Forster (2001) explore modelling health as an investment good; Carbone et al. (2005) allow for individual adaptation to an anchoring health state. These features are not essential in this analysis, hence neglected for parsimony.

$$\sum_{j=1}^{J} P_{jt} E_{jt} \leq y_t + w_t L_t$$
$$\sum_{j=1}^{J} \tau_{jt} E_{jt} = T - L_t \qquad j=1,..J$$

The total expenditure in commodities belonging to the effort vector needs to be met by exogenous income (y_t) and labour income ($w_t L_t$): for simplicity it is assumed that effort choices do not impact on wages and working hours. The amount of time needed to consume a unit of commodity E_{jt} is denoted τ_{jt} ; the total time available, net of working hours, needs therefore to equal the time required for consumption.

The transition equation (2) ensures a recursive nature for this maximisation problem; its Bellman equation is therefore:

$$V = \max_{E_t} U\left(E_t, H_t\right) + EV\left(H_{t+1}\right)$$

The solution of this maximisation problem[13] consists of the demands for health and for each of the effort factors:

$$\begin{aligned} H &= f_h\left(C, \mu\right) \\ E_j &= f_j\left(C, \mu\right) \end{aligned} \qquad (3)$$

The assumption that health outcomes are monotonically increasing in effort, made in the Roemer model, remains sensible in this behavioural framework: healthy lifestyles and education in general are believed to improve health; however, individual preferences, β and π_t may dictate a utility maximising behaviour which diverges from the simple intertemporal maximisation of the health stock.

In section 5, the system consisting of this set of demand equations is estimated, unveiling the triangular relationship between circumstances, effort and health.

3. Data

3.1 The National Child Development Study (NCDS)

The NCDS follows the cohort of nearly 17,000 individuals born in Great Britain in the week of 3rd March 1958. Individuals are followed from birth to the age of 46. Parents were interviewed for the first time in 1958; extensive medical data on children was collected together with comprehensive information about the socioeconomic characteristics and educational achievements of their parents. Posterior interviews were conducted in 1965, 1969, 1974, 1981, 1991, 1999 / 2000 and 2004. Information in the first three waves of the survey was obtained from parents and school teachers. At the age 7 and 11, ability tests were administrated in mathematics and reading. In 1978, the public examination results were made available from schools and added to the NCDS.

[13] Corner solutions are ruled-out by assumption.

During this period of childhood and adolescence, data on some aspects of parental health was systematically collected, such as incidence of hereditary conditions in the family. Parental occupation and education, exposure to financial difficulties and other socioeconomic characteristics of the household were also recorded in these first three waves of the survey[14].

Questionnaires from waves 4 to 7 were addressed to cohort members (rather than their parents) and cover a broad range of subjects grouped in the following categories: employment; income; health and health-related behaviour; citizenship and values; relationships; parenting and housing; education and training.

The issue of attrition has been considered both in research papers and in reports produced by the NCDS advisory panel. Attrition does not seem to be associated with socioeconomic status, as shown in Case et al. (2005), and has modest positive correlation with cohort members' unemployment as reported by Lindeboom et al. (2006). In this paper, a variable addition test was carried-out to investigate whether attrition is a problem when the main outcome of interest is health: ordered probit regressions were used to ascribe whether being in subsequent waves of the panel is a determinant of health status. The results show that, controlling for the set of regressors used in Section 5, the fact that an individual is observed in subsequent waves of the NCDS is not significantly associated with his self-assessed health.

3.2 Variables: health, circumstances and effort

The main health outcome considered in this paper is self-assessed health (SAH) measured in a four-point scale: excellent, good, fair and poor health[15]. SAH is measured when the cohort members are 23, 33, 42 and 46 years old: together with more objective measures of health obtained during the cohort members' childhood and adolescence, this gives a general account of the trajectory of individual health from birth to middle-age. SAH is widely used in health economics and was shown to predict mortality and deterioration of health even after controlling for the medical assessment of health conditions: Idler and Kasl (1995) provide an extensive literature review on this issue. In the specific case of the NCDS, the focus on SAH is also corroborated by its high correlation with reported disability and number of hospitalisations[16].

Two sorts of circumstance variables are considered: the parental socioeconomic background of the cohort members and their congenital and childhood health conditions.

The socioeconomic background of the cohort members is characterised by a comprehensive set of variables. The NCDS allows us to trace the social class of the parents and of both grandfathers of the cohort members. This is derived from the respective Registrar General's Social Class in the first three waves of the survey (for parents) and at the time in which parents left school (for the grandfathers). Following the literature on the NCDS, data on wages was not taken directly into account given

[14] The socioeconomic characterisation of the household includes information about the occupation of the grandparents of the cohort members, whether or not they lived in the household.
[15] In the latest wave of the survey, SAH is however measured in a five-point scale which also includes the category of "very poor health".
[16] See Case et al. (2005, pp. 370).

substantial non-response. Along the lines of Case et al. (2005) and Lindeboom et al. (2006), this was replaced by the incidence of financial difficulties during the childhood of the cohort members. The number of years of schooling of the mother and of the father is also included in the set of circumstances.

The proxies for health endowment used in this paper have all been cited in the literature as systematic determinants of adult health. Birthweight is taken as the main indicator of health at birth; dummy variables for whether the mother smoked after the fourth month of pregnancy and for whether the child was breastfed are included as controls. The NCDS provides information about a comprehensive set of morbidities experienced by the child up until the age of 16. Measures of morbidity, which aggregate 12 categories of health conditions, are constructed according to Power and Peckham (1987) and treated as circumstances. Dummy variables for the occurrence of chronic diseases in the parents and for the incidence of hereditary conditions such as diabetes and epilepsy in parents, brothers and sisters of the cohort members complement the information on health endowments. Dummy variables for whether the child was obese at the age of 11 and 16 and for whether both parents were smokers in 1974 are also treated as circumstances.

The effort factors considered in the paper are, on the one hand, health-related lifestyles such as cigarette smoking, alcohol consumption, consumption of fried food and, on the other hand, educational attainment: these are partially determined by circumstances, but also reflect individual choices.

All the variables used to proxy lifestyles are based on self-reported information. The variable for cigarette smoking is the self-reported number of cigarettes smoked per day. Alcohol consumption is measured by the number of units of alcohol consumed on average per week: NCDS respondents are asked about their weekly consumption of a wide range of alcoholic drinks (glasses of wine, pints of beer and so forth). These were then converted to units of alcohol using the UK National Health Service official guidelines[17]. Educational attainment is measured by the highest academic qualification awarded to cohort members. The summary statistics of the main variables used in the paper is shown in Table 1.

4. Testing and measuring inequality of opportunity in health

The existence of inequality of opportunity in health can be tested using the set of conditions defined in section 2.2. As explained above, the data are consistent with inequality of opportunity if $\forall c \neq c', \forall e, F(.|c,e) \succ_{FSD} F(.|c',e)$. In order to illustrate the application of this condition to the NCDS data, three social types are defined on the sole basis of the social class of the cohort members' father in 1974: a top class including professional and managerial workers, a middle class including partially skilled non-manual and skilled manual workers, and a bottom class including unskilled manual and unemployed workers. This particular classification, adopted for simplicity, is however not essential for this section's results.

[17] These are publicly available at: http://www.nhsdirect.nhs.uk/magazine/interactive/drinking/index.aspx .

The outcome of interest is self-assessed health at age 46, measured in a five-point scale, which constitutes a common discrete support to all types. Given the existence of a common discrete support, Kolmogorov-Smirnov test procedures were carried-out to test for first degree stochastic dominance between types; this approach was previously used in the literature by Lefranc et al. (2006) and Devaux et al. (2008). Table 2 shows the results of these tests: the distribution of health in the top social class dominates at first degree that of the middle class which, in turn, dominates, also at first degree, the outcome distribution of the bottom social type at the 5% significance level. These results establish the existence of inequality of opportunity between types.

In Section 2, two alternative approaches to the measurement of inequality of opportunity were presented. The first of them, the Gini-opportunity approach, is implemented here using the social types defined for testing for stochastic dominance. The first column of Table 3 shows the value of the Gini-opportunity index for health, in the four latest waves of the NCDS; this index measures the extent of inequality of opportunity between the three social types when the cohort members were 23, 33, 42 and 46 years old. To avoid sampling error, the standard errors of the Gini-opportunity indices are bootstrapped in each wave, with independent re-sampling within each of the three types.

The second column of Table 3 presents the values of the indirectly standardised pseudo-Gini coefficient $G(\hat{h}_i)$, which measures the overall inequality that is attributable to circumstances, avoiding the subjective definition of social types. It is computed according with the egalitarian-equivalence approach described in Section 2. The circumstances used in the standardising regression are the following[18]: gender, socioeconomic status of the father and of both grandfathers, number of years of education of the father and of the mother, indicators for whether the father and the mother were smokers in 1974, birthweight, incidence of physical and mental impairments during childhood and adolescence, exposure to financial hardship at age 11 and at age 16, indicators for the prevalence of diabetes, epilepsy and other (unspecified) chronic conditions in the family and a dummy variable for whether the cohort member was obese at age 16.

The third column of Table 3 displays the values of the conventional health pseudo-Gini coefficient $G(h_i)$. As seen in Section 2, this measure treats all the sources of variation in health as circumstances. Under this extreme assumption, there is no distinction between inequality of opportunity and inequality of outcomes: $G(h_i)$ is therefore an upper bound to the extent of inequality of opportunity.

These inequality measures rely on very different assumptions, hence they do not necessarily bring about the same ranking of social states. All the indices show, in this case, an increasing trend, as the 1958 cohort ages and the prevalence of illness mounts. In the case of the Gini-opportunity index, this indicates that the long term association between parental socioeconomic status and the cohort members' health is far from being restricted to childhood and adolescence. The values of $G(\hat{h}_i)$ exhibit a similarly increasing trend, indicating that the set of childhood circumstances used in its computation constitute an increasingly prominent cause of inequality of opportunity in

[18] As explained above, this standardisation procedure is in line with van Doorslaer et al. (2004), in the sense that only circumstance variables are used as standardising variables in the first stage regression.

health. $G(\hat{h}_i)$ is, as seen above, a lower bound for the inequality of opportunity in health. The results of Table 3 show that the gap between this lower bound and $G(h_i)$, the upper bound for inequality of opportunity, is considerable in every wave. This makes clear that the focus on inequality of opportunity, instead of on inequality of outcomes, matters not only in terms of the underlying normative theories, but also with respect to the final conclusions attained.

The measures implemented in this section quantify the extent of inequality of opportunity in health, but fail to illuminate the channels by which the uneven circumstances affect health outcomes. The structural model estimation is therefore crucial to shed light over this complex network of effects.

5. Estimation results: reduced form and structural model

The behavioural model presented in Section 2.4 is estimated in two stages. First, the reduced form is considered in order to assess the association between circumstances and the health at age 46. In a second stage, the structural form, consisting of a system of demand equations for health and for each of the effort factors, is estimated by full information maximum likelihood, allowing for the system errors to be freely correlated and relying on exclusion restrictions to ensure identification. On the one hand this strategy accounts for individual heterogeneity, unobserved to the researcher although not necessarily to the cohort members, unveiling the true causal relationships at stake. On the other hand, the system coefficient estimates illuminate the channels by which circumstances influence health, which are crucial to infer policy implications.

5.1 Reduced form estimation

The first column of Table 4 shows the results of the ordered probit regression of the categorical self-assessed health variable[19] (SAH) on circumstances. A general-to-simple *kitchen sink* approach was followed, starting with a large number of regressors, all of them potential circumstances. The first point of interest concerns the statistically significant coefficients of the dummy variables corresponding to the social class of the father of the cohort members: having as term of comparison the bottom social class, the effect of having a father or male head of household in the middle occupational category is positive and very significant. This positive effect is reinforced in the case of individuals whose father belongs to the top social class. These facts mirror the results of the stochastic dominance analysis of the previous section, establishing once more the existence of inequality of opportunity for health in the NCDS cohort.

The number of years of education of the mother is significantly and positively associated with a better health status in all the attempted specifications of the model[20]. The converse happens with the education of the father, which proves to be insignificant after controlling for the father's social class. This result is in line with Case et al.'s

[19] As seen above, this is measured in a five-point scale ranging from 1 (very poor health) to 5 (excellent health).
[20] This is true even when the social class of the mother is included as a regressor. This variable is however highly collinear with the occupational status of the father or male head of household. As shown in Table 4, the social status of grandfathers proves to be statistically insignificant after controlling for parental background.

(2005, pp 377) work with the NCDS: while the mother's education seems to be an important circumstance in itself, the education of the male head of household matters only as a determinant of income and social position; once these are controlled for, its statistical significance disappears.

Financial difficulties affecting the cohort member's household at age 16, reduce the probability of reporting a satisfactory health state in middle age. This association agrees with recent evidence reported by Propper et al. (2004) showing that spells of low income in early years affect health in childhood and adolescence. According to the results of Table 4, this association is also persistent, since it remains statistically significant in adulthood.

Health endowments also matter: the incidence of illness in adolescence is significantly correlated with a deterioration of self-reported health at age 46. Obesity at age 16 is also highly correlated with a deterioration of adult health. However, the observed inherited health conditions are not statistically significant determinants of SAH, although the respective coefficients have the expected sign.

Estimating the reduced-form model of the first column of Table 4 allows one to illustrate the global impact of circumstances on SAH at age 46. However, this model omits important determinants of health which, in the present context, constitute effort factors. These are added to the circumstances in the model of the second column of Table 4. This paper uses data on individual lifestyles at age 33: instead of current lifestyles, these are employed to rule-out reverse causality of the health status on behavioural choices. When this comprehensive set[21] of potentially relevant effort factors is added to the model, the qualitative impact of circumstances stays constant; at the same time, the magnitude of the individual coefficients is, as expected, reduced.

After controlling for many of the factors that individuals presumably control, and including among them educational attainment and own social class, circumstances are still strongly associated with the health outcomes; this corroborates the existence of significant inequality of opportunity. It is also salient the fact that birthweight remains statistically significant after controlling for ill health conditions at age 16. This is in accord with the *fetal-origin hypothesis* and *life course* models but not with *pathway* models such as Marmot et al. (2001): clearly the circumstances in early adulthood do not capture everything that matters for predicting adult health.

Amongst effort factors, the detrimental effect of cigarette smoking is prominent. This is in line with most of the literature: Power and Peckham (1987), Marmot et al. (2001), Contoyannis and Jones (2004) and Balia and Jones (2007) report similar results. Along the same lines, obesity at the age of 16 remains statistically significant after controlling for a set of dietary choices; among these, only the consumption of fried food is significantly associated[22] with health in adulthood. This suggests that circumstances

[21] The comprehensiveness of this set of controls and the correct specification of the model is corroborated by the RESET test statistic: Chi2(1) = 1.30, with a P value of 0.275.

[22] The weekly consumption of alcohol is never statistically significant in these models and very weakly correlated with health. This result is robust to the specification of a quadratic form aimed at capturing any evidence of beneficial effects of moderate drinking. This is in line with the literature which shows that, absent cases of alcoholism, the detrimental effects of alcohol consumption tend to become apparent after the age of 46. However, under-reporting is also a plausible explanation for this lack of correlation.

have a direct effect on health outcomes which is beyond their influence on lifestyles. Structural form estimates are therefore essential to illuminate this net of direct and indirect effects. On top of this, the estimation of a structural system of equations allows modelling unobserved heterogeneity, hence to identify causal links between circumstances, effort and health.

5.2 Structural form

5.2.1 Identification

The structural model presented in section 2.4 consists of a recursive system of equations which models jointly the demand for health and for each of the effort factors. The first equation of the model is the demand for health. This is modelled as a function of a set of circumstances and of three effort factors: cigarette smoking and consumption of fried food, which were statistically significant determinants of health in the models of last section, and educational attainment. These effort factors are, in turn, modelled in the other equations of the system.

The second and third equations of the system concern cigarette smoking. The data on the number of cigarettes smoked per day shows a spike at zero, which is a typical feature of cigarette smoking data. In order to take this into account, two equations are estimated: the first is a probit model, estimated for the whole sample, for whether an individual is a smoker or a non-smoker; the second equation has the logarithm of the number of cigarettes smoked (continuous variable) as the dependent variable and is estimated only for the individuals who are smokers.

The fourth equation of the system is an ordered probit equation featuring the weekly consumption of fried food as the dependent variable. Avoiding this type of food is a healthy lifestyle, hence this variable takes progressively higher values as the detrimental diet is avoided[23].

The fifth, sixth and seventh equations of the system constitute a generalised conditional probit for the highest academic qualification attained. This specification allows estimating the impact of circumstances on the probability of attaining each educational level. The dependent variables of these three equations are cumulative dummies for the highest academic qualification attained: academic degree or equivalent, A-levels or higher and O-levels or higher; the omitted category corresponds to the absence of secondary education qualifications.

The identification strategy for this recursive system is the standard exclusion restrictions approach suggested by Maddala (1983). The identification restriction in this case is that at least one variable in the effort factor equations is not included in the health equation.

The exclusion restriction for cigarette smoking is an indicator variable for whether cohort members had arguments with parents about the risks of smoking at age 16. This

[23] The variable takes the value 1 if fried food is consumed everyday more than once a day, the value 2 if it is consumed once a day, the value 3 if it is consumed 3 to 6 days a week, the value 4 if fried food is consumed 2 to 3 times a week, the value 5 if it is consumed less than once per week and the value 6 if respondents declare that fried food is never part of their diet.

is a very good predictor of individuals becoming smokers at age 33, but should not influence health in adulthood by any other means than that.

The equation for the consumption of fried food is over-identified by three exclusion restrictions: weekly consumption of vegetables, alcohol and sweets at age 33, which are strongly correlated to the consumption of fried food, but, as patent in Table 4, do not have a statistically significant association with health at age 46. To assess statistically the over-identification of this equation, the three exclusion restrictions were jointly used to instrument the consumption of fried food in the health equation; the value of the Hansen-Sargan over-identification test statistic does not allow the rejection of the null at any significance level[24].

The educational attainment equations are identified using the scores of mathematics tests answered by students at the age of 11; these have been shown to be very good predictors of overall educational attainment but, as seen in Section 5.1, they do not influence health by other relevant channels than education itself.

5.2.2 Estimation strategy

The system is estimated by full information maximum likelihood, allowing the system errors to be freely correlated in order to take into account unobserved common factors that impact both the demands for health and for the effort factors. This way of dealing with selection on unobservables has been followed in recent papers such as Deb and Trivedi (2006), Vera Hernandez (2003) and Pudney and Shields (2000). Since the computation of multidimensional integrals is required, a maximum simulated likelihood procedure is implemented using the Geweke-Hajivassiliou-Keane (GHK) simulator.

Table 5 shows the estimates of the reciprocal correlations between the system errors. The correlation between the errors of the health equation and of the diet equation is positive and statistically significant: this suggests that unobserved factors influence positively both the probability of reporting good health and the probability of choosing a diet that avoids fried food. In addition, the results indicate that unobserved third factors that increase the probability of avoiding fried food, also reduce the probability of an individual being a smoker. This is in line with the evidence that suggests the existence of an individual (genetic or not) propensity for addictive behaviours, affecting simultaneously cigarette smoking and the consumption of calorific food. The table correlations also suggest the existence of unobserved common factors influencing positively the probability of choosing a healthy diet and the probability of having a university degree. The correlation between the different levels of educational attainment and between the probability of being a smoker and the number of cigarettes smoked per day is also statistically significant. The estimates of the univariate models of Section 5.1 fail to take all these effects into account.

[24] The value of the Hansen "J" Statistic was in this case 1.65, corresponding to a P value of 0.4306 hence corroborating the over-identification of this equation. Also the Kleibergen-Paap test allows the rejection of under-identification when the three excluded variables are use to instrument the consumption of fried food in the health equation.

5.2.5 Results

Table 6 shows the system estimation results. The health equation estimates give a general picture which is similar, in most cases, to the one given by the univariate models: a higher socioeconomic status of the father has a statistically significant and positive effect on health; the incidence of illness in childhood and adolescence has a negative effect on SAH; obesity keeps having a statistically significant negative effect on health; own socioeconomics status at age 33 remains positively associated with better health at age 46 and cigarette smoking remains significantly detrimental. An important exception is the fact that the consumption of fried food is no longer a statistically significant determinant of health. This can be explained by the fact that, as seen in Table 5, there are unobserved common factors that affect both the health equation and the equation for the consumption of fried food. The system estimates take into account these effects, making clear that the association present in the univariate model is not a true causal link[25].

The second and the third equations of the system concern cigarette smoking. The probability of an individual being a smoker at the age 33 does not seem, in general, to be determined by the social class of the father. However, having experienced financial difficulties in childhood, has a positive and statistically significant effect on the probability of an individual being a smoker. Parental smoking and the occurrence of arguments with the parents of the cohort members about the risks of smoking are statistically significant predictors of smoking in early adulthood. Also interesting is the statistically significant negative effect of the incidence of chronic diseases in the family, other than diabetes and epilepsy, on smoking. This is in line with the hypothesis that physical frailty, perceived by the individual but not necessarily by the researcher, may lead individuals to adopt healthy lifestyles in order to offset health risks[26]. Finally, the results indicate a clear socioeconomic and educational gradient in the probability of smoking: those with higher qualifications are less likely to smoke, even after controlling for own and parental socioeconomic status. This result has policy implications, for it suggests that, in order to reduce health inequalities, complementary policies may be required, for example, in the educational sector.

The evidence concerning the number of cigarettes smoked per day is mixed: there is neither a clear socioeconomic gradient nor an educational gradient. This is in accord with papers such as Jones (1989), which show that cigarette smoking is best modelled by a hurdle specification: education and social status reduce the probability of an individual becoming a smoker; however, for the individuals who are already smokers, tobacco is a normal good.

In the fourth equation of the system, obesity at age 16 has a positive and statistically significant effect on the avoidance of fried food. This is in line with the aforementioned

[25] It is unlikely that a weak instruments problem is interfering with this result. The equation for the consumption of fried food is over-identified by three exclusion restrictions. In order to assess the strength of these restrictions, they were used to instrument the consumption of fried food in the health equation. A test of weak instruments was then carried-out: the value of the F test statistic is 55.084, well above any of the Stock-Yogo critical values for weak instruments. Although the Stock-Yogo critical values are tabulated under a set of restrictive assumptions, the test results indicate a high statistical correlation between the instruments and the consumption of fried food.

[26] This thesis is corroborated by the recent findings of Balia and Jones (2007), based on the British Health and Lifestyle Survey.

rationale of individual offsetting of health risks in face of perceived frailty. More importantly, it becomes apparent that the harmful impact of child obesity on adult health is a direct one, which does not operate solely through dietary choices in adulthood. This suggests that childhood obesity should be treated as policy objective in itself.

A dummy variable for whether the cohort members' diet was prescribed by a doctor is included in the equation to, once more, discard the possibility of reverse causality of health on dietary choices; this variable is however statistically insignificant. Educational qualifications also matter as does the cohort members' socioeconomic status, although there is no evidence of a gradient.

Finally, the educational achievement equations (fifth, sixth and seventh equations of the system) constitute a generalised conditional probit. Parental social position and education have a positive and statistically significant effect of on the highest academic qualification attained by the cohort members. Experiencing financial difficulties in childhood has a negative effect, which is statistically significant only in the equation for the probability of reporting O-levels or equivalent versus no secondary education qualification. Low birthweight and the incidence of disease at the age 16 also have a negative impact on educational outcomes. This exemplifies how circumstances produce effects on health through their influence on the exertion of (educative) effort.

A final point concerns the estimation of this system of equations for separate sub-samples of men and women. As made clear by the results in Table 7, the main results of this section extend to the case in which the system is estimated only for women. This is however not the case when the system is estimated only for men: the parental social class variables are not statistically significant in the health equation suggesting that inequality of opportunity tends to be stronger for women than it is for men.

6. Conclusions

The paper provides evidence indicating the existence of inequality of opportunity in heath among NCDS cohort members of different parental background. It puts forward two alternative approaches to measuring it: the results suggest that the focus on inequality of opportunity (instead of inequality of outcomes) is likely to significantly reduce the magnitude of the health inequalities that are deemed illegitimate.

The paper then turns from identifying and measuring inequality of opportunity, to explaining it. Two types of results are made available through the estimation of a structural model. The reduced form results show that parental social status, the education of the mother (but not of the father), and the exposure to financial difficulties during childhood and adolescence are statistically significant determinants of health in adulthood, hence relevant circumstances. Health endowments are also important: ill health in early age and childhood obesity are negatively associated with health at age 46. These results are in line with most of the literature on the long term impact of childhood circumstances.

The illumination of the causal channels through which circumstances affect health outcomes requires the estimation of the structural form of the model. These estimates confirm the importance of circumstances as determinants of health in adulthood. They also show that each effort factor is impacted by a different set of circumstances: for example, parental social class influences significantly the demand for education but fails to explain cigarette smoking; this, in turn, is affected mainly by parental smoking and health endowments such as the perceived prevalence of chronic diseases in the family.

These results suggest clear policy implications. Some unjust circumstances appear to be amenable to policy only during childhood: childhood obesity for example, has a clear direct impact on health in adulthood, which is beyond its effect over dietary choices. Furthermore, since the influence of circumstances on health is often channelled through effort, some key policies to reduce inequality of opportunity in health may be found outside the health care system. The influence of education on lifestyles suggests that complementary educational policies may be of paramount importance to reduce health inequalities.

References

Arneson, R. 1989. Equality and equal opportunity for welfare. Philosophical Studies 56, 77-93.

Balia, S and Jones, A. M. 2008. Mortality, lifestyle and socio-economic status. Journal of Health Economics 27(1), 1-26.

Barker, D.J.P. 1995. Foetal origins of coronary disease. British Medical Journal 311, 171-174.

Bensaid, J. and Fleurbaey, M. 2003. La fiscalité des ensembles de budgets: Quelques Propositions. Economie et Prevision 22, 110-111.

Betts, J. and Roemer, J. 1998. Equalizing opportunities through educational finance reform. U. C. Davis Working Paper No. 99-08.

Bourguignon, F. Ferreira, F and Menéndez, M. 2005. Inequality of Opportunity in Brazil. World Bank Policy Research Working Paper Series No. 3174.

Case, A. Ferting, A. and Paxon, C. 2005. The lasting impact of childhood health and circumstance. Journal of Health Economics 24, 365-389.

Carbone, J. C., Kverndokk, S. and Rogeberg, O. J. 2005. Smoking, health, risk and perception. Journal of Health Economics 24(4), 631-653.

Cohen G. A. 1989. On the currency of egalitarian justice. Ethics 99, 906-944.

Contoyannis, P. and Jones, A.M. 2004. Socio-economic status, health and lifestyle. Journal of Health Economics 23(5), 965-995.

Currie, J. Stabile, M. 2004. Socioeconomic status and health: why is the relationship stronger for older children? American Economic Review 93(5), 1813-1823.

Dardononi, V., Fields, G., Roemer, J. and Sanchez, L. 2005. How demanding should equality of opportunity be and how much have we achieved. In Gary Fields, David Grusky and Steven Morgan (eds.), Mobility and Inequality: Frontiers of Research from Sociology and Economics, Stanford University Press.

Dardanoni, V. and Wagstaff, A, 1990. Uncertainty and the demand for medical care. Journal of Health Economics 9(1), 23-38.

Deb, P and Trivedi, P. 2006. Specification and Simulated Likelihood Estimation of a Non-normal Outcome Model with Selection: Application to Health Care Utilization. Econometrics Journal, 9(2) 307-331.

Devaux M., Jusot F., Trannoy A. and Tubeuf S. 2008. La santé des séniors selon leur origine sociale et la longévité de leurs parents. Economie et Statistique, in press.

Dworkin, R. 1981. What is equality? Part 2: Equality of resources. Philosophy & Public Affairs 10, 283-345.

Fleurbaey, M. and Schokkaert, E. 2007. Unfair inequalities in health and health care. CORE Discussion Paper 2007/90.

Forster, M. 2001. The meaning of death: some simulations of a model of healthy and unhealthy consumption. Journal of Health Economics 20, 613-638.

Gravelle, H. 2003. Measuring income related inequality in health: standardisation and the partial concentration index. Health Economics 12, 803-819.

Grossman, M. 1972. On the concept of health capital and the demand for health. Journal of Political Economy 80(2), 223-255.

Jones, A. M. 1989. A double hurdle model of cigarette consumption. Journal of Applied Econometrics 4, 23-39.

Hadar, J. and Russell, W. R. 1969. Rules for Ordering Uncertain Prospects, American Economic Review 59(1), 25-34.

Hanoch, G. and Levy, H. 1969. The efficiency analysis of choices involving risk. Review of Economic Studies 38, 335-346.

Kenkel, S. D. 1995. Should you eat breakfast? Estimates from health production functions. Health Economics 5, 15-25.

Kuh, D. J. and Wadsworth, M. E. 1993. Physical health status at 36 years in a British national cohort. Social Science and Medicine 37(7), 905-916.

Lambert, P. J. 2001. The Distribution and Redistribution of Income. Manchester University Press.

Lechene, V. and Adda, J. 2004. On the Identification of the effect of smoking on mortality. CeMMAP working paper series No. CWP13/04.

Lefranc, A., Trannoy, A., and Pistolesi, N. 2005. Inequality of opportunities vs inequality of outcomes: Are western societies all alike? ECINEQ working paper series No. 54.

Lefranc, A., Trannoy, A. and Pistolesi, N. 2006. Equality of opportunity: Definitions and testable conditions with an application to income in France. THEMA working papers series No. 2006-13, University of Cergy-Pontoise.

Le Grand, J. and Burchardt, T. 2002. Constraint and Opportunity: Identifying Voluntary Unemployment. CASE discussion paper series No. 55, LSE.

Lindeboom, M., Llena-Nozal, A. and van der Klaauw, B. 2006. Disabiliy and work: the role of childhood circumstances. Timbergen Institute discussion paper 06-039/3, Tinbergen Institute.

Levy, H. 2000. Stochastic Dominance: Investment Decision Making Under Uncertainty. Kluwer Academic Publishers

Madala, G. S. 1983. Limited dependent and qualitative variables in econometrics. Cambridge University Press.

Marmot, M.G., Brunner, S., Hemingway, S. 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, 301-307.

Mullahy, J. and Portney, P. 1990. Air pollution, cigarette smoking and the production of respiratory health. Journal of Health Economics 9, 193-205.

Power, C. and Peckham, C. 1987. Childhood morbidity and adult ill-health. National Child Development Study User Support Working Paper No. 37.

Propper, C., Burgess, S. and Rigg, J.A. 2004. The Impact of Low Income on Child Health: Evidence from a Birth Cohort Study. CASE discussion paper series No. 085, LSE.

Pudney, S. and Shields, M. A. 2000. Gender and racial discrimination in pay and promotion for NHS nurses. Oxford Bulletin of Economics and Statistics 62, 801-35.

Ravelli, A.C.J., van der Meulen, J.H.P., Michels, R.P.J., Osmond, C., Barker, D.J.P., Hales, C.N., Bleker, O.P. 1998. Glucose tolerance in adults after prenatal exposure to famine. The Lancet 351, 173-177.

Rosa Dias, P and Jones, A. M. 2007. Giving equality of opportunity a fair innings Health Economics 16, 109-112.

Roemer J. E. 1998. Equality of Opportunity. Harvard University Press.

Roemer J. E. 2002. Equality of opportunity: A progress report. Social Choice and Welfare 19, 455-471.

Rothschild, M. and Stiglitz, J. 1970. Increasing Risk: A Definition. Journal of Economic Theory, 225-243.

Sen, A.K. and Foster, J.E. 1997. On Economic Inequality: Expanded edition. Oxford University Press.

Van Doorslaer, E. and Jones, A. M. 2003. Inequalities in self-reported health: validation of a new approach to measurement. Journal of Health Economics 22, 61-87.

Van Doorslaer, E. and Koolman, X. 2004. Explaining the differences in income-related health inequalities across European countries. Health Economics 13, 609-628.

Vera-Hernandez, M. 2003. Structural estimation of a principal agent model: moral hazard in medical insurance. The RAND Journal of Economics 34, 670-93.

Table 1: Summary statistics

Full sample

Variable	Mean	Std. Dev.	Min	Max
Self-assessed health, age 46	3.987719	0.9302554	1	5
Male	0.5171652	0.4997187	0	1
Parental socioeconomic status at birth: high	0.2727015	0.4453612	0	1
Parental socioeconomic status at birth: middle	0.49983	0.5000141	0	1
Paternal grandfather's socioeconomic status	1.975576	0.7470104	1	3
Maternal grandfather's socioeconomic status	2.04248	0.7366398	1	3
Years of education: father	9.904075	1.621967	7	16
Years of education: mother	9.916638	1.376012	7	16
Indicator: mother smoker, age 16	0.7865378	1.010508	0	4
Indicator: father smoker, age 16	1.119048	1.136957	0	4
Indicator: maternal smoking after 4th month of pregnancy	0.3364165	0.472497	0	1
Indicator: breastfead	0.6421394	0.4793864	0	1
Birthweight	128.3177	72.43585	0	509
Physical / mental impairments, age 16	2.236591	1.541278	0	10
Indicator: financial hardship, age 11	0.0714425	0.2575708	0	1
Indicator: financial hardship, age 16	0.0789546	0.269677	0	1
Indicator: diabetes in parents, brothers or sisters	0.0212642	0.1442684	0	1
Indicator: epilepsy in parents, brothers or sisters	0.073906	0.2616263	0	1
Indicator: other hereditary chronic condition	0.025154	0.1565977	0	1
Indicator: chronic conditions in cohort member's mother, age 16	0.0477003	0.2131386	0	1
Indicator: obesity, age 16	0.0324388	0.1771673	0	1
Indicator: university degree or equivalent	0.2313824	0.4217384	0	1
Indicator: A-levels or higher qualification	0.3206419	0.4667478	0	1
Indicator: O-levels, or higher qualification	0.8212712	0.3831451	0	1
Mathematics test score, age 11 (scores range from 0 to 40)	15.23885	11.01308	0	40
Indicator : smoker, age 33	0.3197992	0.4664195	0	1
Number of cigarettes per day	5.543246	9.519264	0	70
Arguments with parents about risks of smoking	0.0913892	0.2881695	0	1
Avoidance of fried food in diet: weekly frequency (1 to 6), age 33	4.538137	0.9861445	1	6
Consumption of vegetables, age 33	0.6580174	0.638489	0	2
Weekly consumption of alcohol, age 33	2.453389	1.619937	0	4
Consumption of sweets: weekly frequency, age33	4.152178	1.667634	1	9
Socioeconomic status: high (age 33)	0.5977131	0.4903824	0	1
Socioeconomic status: middle (age 33)	0.2081837	0.4060281	0	1

Estimation sample used in Section 5 (No. observations: 4442)

Variable	Mean	Std. Dev.	Min	Max
Self-assessed health, age 46	4.019153	0.9071588	1	5
Male	0.4833258	0.4997782	0	1
Parental socioeconomic status at birth: high	0.3017125	0.4590529	0	1
Parental socioeconomic status at birth: middle	0.5072105	0.5000043	0	1
Paternal grandfather's socioeconomic status	1.890102	0.7379079	1	3
Maternal grandfather's socioeconomic status	1.994581	0.7376612	1	3
Years of education: father	9.968004	1.626594	7	16
Years of education: mother	9.996845	1.377723	7	16
Indicator: mother smoker, age 16	0.7178909	0.9783318	0	4
Indicator: father smoker, age 16	1.043713	1.118162	0	4
Indicator: maternal smoking after 4th month of pregnancy	0.3016125	0.4590101	0	1
Indicator: breastfead	0.6885985	0.4631187	0	1
Birthweight	127.4932	64.46216	0	509

Physical / mental impairments, age 16	2.447724	1.567314	0	9
Indicator: financial hardship, age 11	0.0484452	0.2147293	0	1
Indicator: financial hardship, age 16	0.0464173	0.2104108	0	1
Indicator: diabetes in parents, brothers or sisters	0.022082	0.1469669	0	1
Indicator: epilepsy in parents, brothers or sisters	0.0687247	0.2530138	0	1
Indicator: other hereditary chronic condition	0.0236593	0.1520025	0	1
Indicator: chronic conditions in cohort member's mother, age 16	0.0434881	0.203976	0	1
Indicator: obesity, age 16	0.0466426	0.210896	0	1
Indicator: university degree or equivalent	0.2323119	0.4223544	0	1
Indicator: A-levels or higher qualification	0.3323569	0.4711112	0	1
Indicator: O-levels, or higher qualification	0.8787742	0.3264263	0	1
Mathematics test score, age 11 (scores range from 0 to 40)	17.6735	10.67408	0	40
Indicator : smoker, age 33	0.2764759	0.4473053	0	1
Number of cigarettes per day	4.567147	8.684014	0	63
Arguments with parents about risks of smoking	0.1191978	0.3240577	0	1
Avoidance of fried food in diet: weekly frequency (1 to 6), age 33	4.604777	0.9495712	1	6
Consumption of vegetables, age 33	0.6547995	0.6314485	0	2
Weekly consumption of alcohol, age 33	1.528616	1.333322	0	4
Consumption of sweets: weekly frequency, age33	4.051769	1.613378	1	9
Socioeconomic status: high (age 33)	0.6336187	0.4818696	0	1
Socioeconomic status: middle (age 33)	0.1962596	0.3972119	0	1

Table 2: Tests for stochastic dominance between types

Null hypothesis	Corrected P value
Null: Type 1 FSD type 2	0.999
Null: Type 1 FSD type 3	0.999
Null: Type 2 FSD type 3	0.959
Null: Type 2 FSD type 1	0.000
Null: Type 3 FSD type 1	0.000
Null: Type 3 FSD type 2	0.051

Table 3: Measures of inequality of opportunity

NCDS wave	Gini-opportunity index	Egalitarian-equivalence approach: $G\left(\hat{h}_i\right)$	Health Pseudo-Gini: $G\left(h_i\right)$
Wave 4: 1981 (age 23)	.0088496 (.0017707)	0.02205	0.10257
Wave 5: 1991 (age 33)	.0165535 (.0015658)	0.02976	0.11304
Wave 6: 1999 / 2000 (age 42)	.018381 (.0018364)	0.03257	0.12765
Wave 7: 2004 (age 46)	.0178522 (.0026443)	0.03380	0.15405

Table 4: Reduced form estimates

Dependent variable: SAH at age 46	Circumstances only		Circumstances and effort	
	Coef.	Std. Err.	Coef.	Std. Err.
Male	.0208803	(0.0326851)	.0392548	(.0427294)
Parental socioeconomic status at birth: high	.2043999	(.0604832)***	.2223055	(.0695089)***
Parental socioeconomic status at birth: middle	.1411163	(.0452732)***	.1056241	(.0521479)**
Paternal grandfather's socioeconomic status	-.0330035	(.0287464)**	.0198249	(.0332016)
Maternal grandfather's socioeconomic status	-.0116359	(.0242637)	-.0082476	(.0277494)
Years of education: father	-.0098132	(.0127866)	-.0091925	(.0143907)
Years of education: mother	.0305556	(.0145409)**	.0249568	(.0165671)
Indicator: mother smoker, age 16	-.0462621	(.0216748)**	-.0424256	(.0252343)*
Indicator: father smoker, age 16	-.0190165	(.0155828)	-.006466	(.0179086)
Indicator: maternal smoking after 4th month of pregnancy	-.0012025	(.044241)	.0318697	(.0510824)
Indicator: breastfead	.0469327	(.0361232)	.0601748	(.041468)
Birthweight	.0003589	(.0002529)	.0006205	(.0003075)**
Physical / mental impairments, age 16	-.0777528	(.0107033)***	-.0729514	(.0123095)***
Indicator: financial hardship, age 11	-.0820376	(.0784684)	-.0349555	(.092093)
Indicator: financial hardship, age 16	-.206553	(.0773989)***	-.1546493	(.0909332)*
Indicator: diabetes in parents, brothers or sisters	-.0849676	(.108588)	-.0824223	(.1229156)
Indicator: epilepsy in parents, brothers or sisters	-.0803568	(.0629973)	-.0711624	(.0739004)
Indicator: other hereditary chronic condition	-.1006471	(.1050638)	-.0944268	(.1223111)
Indicator: chronic conditions in cohort member's mother, age 16	-.0870696	(.0780231)	-.1259423	(.0940664)
Indicator: obesity, age 16	-.2589265	(.0776559)***	-.2655016	(.0875341)***
Indicator: university degree or equivalent			-.0632771	(.0696125)
Indicator: A-levels or higher qualification			.1039904	(.0663569)
Indicator: O-levels, or higher qualification			.0524404	(.062628)
Mathematics test score, age 11 (scores range from 0 to 40)			.0007801	(.0020205)
Indicator (smoker)*Log(No. Cigarettes per day), age 33			-.1218921	(.0159327)***
Avoidance of fried food in diet: weekly frequency (1 to 6), age 33			.0609421	(.0210325)***
Consumption of vegetables, age 33			-.0176862	(.0304288)
Weekly consumption of alcohol, age 33			-.0000392	(.0151047)
Consumption of sweets: weekly frequency, age33			.0029998	(.0116626)
Socioeconomic status: high (age 33)			.1073744	(.0549581)*
Socioeconomic status: middle (age 33)			.1104835	(.0649887)*

Table 5: System errors correlation matrix

Rho	Coef.	Std. Err.	Rho	Coef.	Std. Err.
Health / log(# cigarettes)	-.0319463	(0.0412)	Smoker / fried food	-.1098722	(0.0220)***
Health /smoker	.0285416	(0.0600)	Smoker / degree	.1297514	(0.0454)***
Health / fried food	.0932166	(0.0468)**	Smoker / A-levels or higher	.0425414	(0.0496)
Health / degree	.0118997	(0.0325)	Smoker / O-levels or higher	.1910958	(0.0264)***
Health / A-levels or higher	-.0171162	(0.0317)	Fried food / degree	.09626	(0.0329)***
Health / O-levels or higher	-.0145465	(0.0322)	Fried food / A-levels or higher	.0491199	(0.0303)
Log(# cigarettes) / smoker	-.025113	(0.1552)	Fried food / O-levels or higher	.0342463	(0.0298)
Log(# cigarettes) / fried food	-.0626928	(0.0309)**	Degree / A-levels or higher	.9903971	(0.0039)***
Log(# cigarettes) / degree	.024997	(0.0678)	Degree / O-levels or higher	.9695317	(0.0134)***
Log(# cigarettes) / A-levels or higher	.1142903	(0.0733)	A-levels or higher / O-levels or higher	.9388271	(0.0202)***
Log(# cigarettes) / O-levels or higher	-.2017146	(0.0388)***			

Table 6: Structural form estimates
Number of observations: 4442

	Coef.	Std. Err.
Dependent variable: SAH at age 46		
Parental socioeconomic status at birth: high	.1520387	(.0548274)***
Parental socioeconomic status at birth: middle	.0791978	(.0462509)*
Years of education: mother	.0218121	(.0143421)
Birthweight	.0004253	(.0002435)*
Physical / mental impairments, age 16	-.0742458	(.0110513)***
Indicator: financial hardship, age 11	-.0798622	(.0800013)
Indicator: financial hardship, age 16	-.0768533	(.0776275)
Indicator: other hereditary chronic condition	-.0552032	(.1135692)
Indicator: epilepsy in parents, brothers or sisters	-.0853214	(.0667876)
Indicator: chronic conditions in cohort member's mother, age 16	-.1094562	(.0857164)
Indicator: obesity, age 16	-.2677265	(.0747986)***
Indicator: university degree or equivalent	-.1035845	(.0776368)
Indicator: A-levels or higher qualification	.1439075	(.0702164)**
Indicator: O-levels, or higher qualification	.1158332	(.0744613)
Indicator (smoker)*ln(No. Cigarettes per day), age 33	-.146192	(.0369219)***
Avoidance of fried food in diet: weekly frequency (1 to 6), age 33	-.049113	(.049385)
Socioeconomic status: high (age 33)	.1571853	(.0501965)***
Socioeconomic status: middle (age 33)	.1218061	(.0562687)**
Dependent variable: Smoker (binary indicator)		
Arguments with parents about the risks of smoking, age 16	1,103,539	(.0621943)***
Indicator: mother smoker, age 16	.0495701	(.0223023)**
Indicator: father smoker, age 16	.0834715	(.0205775)***
Parental socioeconomic status at birth: high	-.0150949	(.0680007)
Parental socioeconomic status at birth: middle	-.0015746	(.0565511)
Indicator: university degree or equivalent	-.4640214	(.1174811)***
Indicator: A-levels or higher qualification	.00313	(.1169824)
Indicator: O-levels, or higher qualification	-.6293806	(.07707)***
Socioeconomic status: high (age 33)	-.2211447	(.0596216)***
Socioeconomic status: middle (age 33)	-.0975994	(.0670404)
Years of education: mother	.0822799	(.0178375)***
Physical / mental impairments, age 16	-.0167732	(.0138468)
Indicator: financial hardship, age 11	.1882077	(.0948417)**
Indicator: other hereditary chronic condition	-.3679449	(.154845)**
Indicator: epilepsy in parents, brothers or sisters	-.011569	(.0851059)
Indicator: chronic conditions in cohort member's mother, age 16	.0411411	(.1009445)
_cons	-.862864	(.1894779)***
Dependent variable: Ln(No. Cigarettes per day)	**(Number of observations: 1227)**	
Arguments with parents about the risks of smoking, age 16	.147952	(.0846376)*
Indicator: mother smoker, age 16	.0711708	(.0204446)***
Indicator: father smoker, age 16	.0405411	(.0192028)**
Parental socioeconomic status at birth: high	-.0654537	(.0591578)
Parental socioeconomic status at birth: middle	-.045955	(.0511063)
Indicator: university degree or equivalent	.018523	(.1034411)
Indicator: A-levels or higher qualification	-.2935658	(.0969564)***
Indicator: O-levels, or higher qualification	.2442692	(.081189)***
Socioeconomic status: high (age 33)	-.1072667	(.0553912)*
Socioeconomic status: middle (age 33)	.0629448	(.0612048)
Years of education: mother	-.023988	(.0149075)
Physical / mental impairments, age 16	.007695	(.0125241)
Indicator: financial hardship, age 11	.1201082	(.0789669)
Indicator: other hereditary chronic condition	-.0337527	(.1711175)
Indicator: epilepsy in parents, brothers or sisters	-.0439847	(.0687357)
Indicator: chronic conditions in cohort member's mother, age 16	-.0227499	(.0781784)
_cons	2,630,866	(.2140098)***

Dep. Variable: Avoidance of fried food in diet, age 33		
Indicator: obesity, age 16	.2037492	(.0726168)***
Diet recommended by a doctor, age 33	.0816036	(.1407643)
Indicator: university degree or equivalent	-.138093	(.0817946)*
Indicator: A-levels or higher qualification	.1449821	(.0689828)**
Indicator: O-levels, or higher qualification	.2494402	(.0633997)***
Socioeconomic status: high (age 33)	.1303015	(.0448696)***
Socioeconomic status: middle (age 33)	-.2374111	(.0491029)***
Parental socioeconomic status at birth: high	.0104717	(.0501989)
Parental socioeconomic status at birth: middle	-.0550794	(.0450043)
Years of education: mother	-.0264387	(.0149782)*
Weekly consumption of salads, age 33	.2262754	(.0236651)***
Units of alcohol per week, age 33	-.0976156	(.0121157)***
Weekly consumption of sweets, age 33	.0563589	(.0221102)**
Dep. Variable: University degree or equivalent (indicator)		
Mathematics test score, age 11 (scores range from 0 to 40)	.0441603	(.001925)***
Parental socioeconomic status at birth: high	.261412	(.0689757)***
Parental socioeconomic status at birth: middle	.1254431	(.0583515)**
Years of education: father	.067009	(.0158214)***
Years of education: mother	.1276801	(.0174289)***
Indicator: mother smoker, age 16	-.0128399	(.0211393)
Indicator: father smoker, age 16	-.0589288	(.0199789)***
Indicator: financial hardship, age 11	-.1551059	(.1111651)
Birthweight	-.0002382	(.0003283)
Physical / mental impairments, age 16	-.0276099	(.0128962)**
_cons	-3,608,787	(.1839348)***
Dep. Variable: A-levels or higher qualification (indicator)		
Mathematics test score, age 11 (scores range from 0 to 40)	.0524479	(.0019763)***
Parental socioeconomic status at birth: high	.3428773	(.0695349)***
Parental socioeconomic status at birth: middle	.1478215	(.0606121)**
Years of education: father	.0752321	(.0166349)***
Years of education: mother	.1653061	(.0185223)***
Indicator: mother smoker, age 16	-.0548359	(.0228356)**
Indicator: father smoker, age 16	-.0687658	(.0205546)***
Indicator: financial hardship, age 11	-.1050756	(.1112243)
Birthweight	-.0000825	(.0003311)
Physical / mental impairments, age 16	-.0291884	(.0138813)**
_cons	-3,862,516	(.1992426)***
Dep. Variable: O-levels or higher qualification (indicator)		
Mathematics test score, age 11 (scores range from 0 to 40)	.0463875	(.0022588)***
Parental socioeconomic status at birth: high	.2598224	(.0527071)***
Parental socioeconomic status at birth: middle	.1561767	(.0455699)***
Years of education: father	.0644588	(.0180778)***
Years of education: mother	.1027676	(.0194972)***
Indicator: mother smoker, age 16	-.0947865	(.0195064)***
Indicator: father smoker, age 16	-.0989729	(.0158937)***
Indicator: financial hardship, age 11	-.4979128	(.0857625)***
Birthweight	.0009846	(.0001921)***
Physical / mental impairments, age 16	-.0084091	(.012497)
_cons	-1,116,528	(.2049557)***

Table 7: Structural form estimates: Sample split by gender	Females (No. Obs: 2295)		Males (No. Obs: 2147)	
	Coef.	Std. Err.	Coef.	Std. Err.
Dependent variable: SAH at age 46				
Parental socioeconomic status at birth: high	.1737236	(.0751484)**	.100273	(.0823928)
Parental socioeconomic status at birth: middle	.0991449	(.0635767)	.0267519	(.0701623)
Years of education: mother	.0271209	(.0210395)	.0134766	(.0219768)
Birthweight	.0003362	(.0003331)	.0004736	(.0003784)
Physical / mental impairments, age 16	-.0682339	(.014721)***	-.0763825	(.0169973)***
Indicator: financial hardship, age 11	-.2421496	(.1197724)**	.1170256	(.1130796)
Indicator: financial hardship, age 16	-.0664009	(.1094768)	-.1613189	(.1144146)
Indicator: other hereditary chronic condition	.0119232	(.1730057)	-.0911508	(.1506804)
Indicator: epilepsy in parents, brothers or sisters	-.0854629	(.0965001)	-.1139778	(.0989878)
Indicator: chronic conditions in cohort member's mother, age 16	-.0856005	(.1241235)	-.1794336	(.1248205)
Indicator: obesity, age 16	-.3441254	(.1034829)***	-.1764745	(.1120325)*
Indicator: university degree or equivalent	-.0542937	(.1207724)	-.0105422	(.1167862)
Indicator: A-levels or higher qualification	.0511007	(.1109708)	.1351889	(.1031627)
Indicator: O-levels, or higher qualification	.2068405	(.114249)*	-.0981534	(.1135603)
Indicator (smoker)*ln(No. Cigarettes per day), age 33	-.2169274	(.0459054)***	-.0706518	(.0334612)**
Avoidance of fried food in diet: weekly frequency (1 to 6), age 33	-.0141954	(.0829892)	-.0308025	(.0860439)
Socioeconomic status: high (age 33)	.1043188	(.062544)*	.2466661	(.0844765)***
Socioeconomic status: middle (age 33)	.0392325	(.0978381)	.1748147	(.0800168)**
Dependent variable: Smoker (binary indicator)				
Arguments with parents about the risks of smoking, age 16	1,188,854	(.0923098)***	.8566676	(.0789098)***
Indicator: mother smoker, age 16	.0483321	(.033386)	.0105638	(.0302457)
Indicator: father smoker, age 16	.0939358	(.0290738)***	.0505291	(.0285016)*
Parental socioeconomic status at birth: high	-.0918009	(.0969373)	.1135671	(.0942516)
Parental socioeconomic status at birth: middle	-.05484	(.0799989)	.0738797	(.0794471)
Indicator: university degree or equivalent	-.2257403	(.1536571)	-.3554223	(.1575031)**
Indicator: A-levels or higher qualification	-.3838032	(.1586372)**	.1829644	(.1421127)
Indicator: O-levels, or higher qualification	-.3789566	(.1471448)***	-1,175,395	(.0900648)***
Socioeconomic status: high (age 33)	-.2309016	(.0770357)***	-.163429	(.0962988)*
Socioeconomic status: middle (age 33)	-.0985532	(.1265295)	-.0221651	(.0930138)
Years of education: mother	.0841399	(.0264922)***	.0891548	(.0229481)***
Physical / mental impairments, age 16	-.0109808	(.019578)	-.0359294	(.0195608)*
Indicator: financial hardship, age 11	.1013911	(.1281116)	.1942701	(.1555357)
Indicator: other hereditary chronic condition	-.2572325	(.2167434)	-.5264833	(.218652)**
Indicator: epilepsy in parents, brothers or sisters	-.0704684	(.123052)	.1028565	(.1151883)
Indicator: chronic conditions in cohort member's mother, age 16	-.0777675	(.1618184)	.1638114	(.1241964)
_cons	-.9748284	(.2794675)***	-.5833932	(.2498484)**
Dependent variable: Ln(No. Cigarettes per day)	**No. Obs: 640**		**No. Obs: 587**	
Arguments with parents about the risks of smoking, age 16	.2501684	(.1048293)**	-.6601114	(.0921348)***
Indicator: mother smoker, age 16	.0918863	(.0272007)***	.0213172	(.0350148)
Indicator: father smoker, age 16	.0134491	(.0252834)	.0299491	(.0343476)
Parental socioeconomic status at birth: high	-.0445364	(.0795017)	-.1249169	(.1100626)
Parental socioeconomic status at birth: middle	-.0430467	(.0660178)	-.1068636	(.092451)
Indicator: university degree or equivalent	.0410724	(.1508645)	.1882696	(.1937832)
Indicator: A-levels or higher qualification	-.3182032	(.1315146)**	-.3560098	(.1694554)**
Indicator: O-levels, or higher qualification	.4095827	(.1329145)***	1,368,414	(.1102647)***
Socioeconomic status: high (age 33)	-.1106114	(.0686688)	.111427	(.10807)
Socioeconomic status: middle (age 33)	.0900881	(.1116069)	-.0231093	(.1073839)
Years of education: mother	-.0281842	(.020973)	-.0918979	(.0264061)***
Physical / mental impairments, age 16	-.0015536	(.0162624)	.0585774	(.0226067)**
Indicator: financial hardship, age 11	.1713148	(.1045988)	-.034283	(.1702817)
Indicator: other hereditary chronic condition	-.1277179	(.2122216)	.6696442	(.2636425)**
Indicator: epilepsy in parents, brothers or sisters	-.0580207	(.0992297)	-.1608246	(.1230253)
Indicator: chronic conditions in cohort member's mother, age 16	-.0371704	(.1025634)	-.0474446	(.147968)
_cons	2,446,052	(.2927912)***	3,688,021	(.2894581)***

Dep. Variable: Avoidance of fried food in diet, age 33				
Indicator: obesity, age 16	.1789177	(.0990026)*	.2609556	(.1143207)**
Diet recommended by a doctor, age 33	.0159972	(.1980767)	.1461485	(.1843482)
Indicator: university degree or equivalent	-.35298	(.1576139)**	.2935587	(.104031)***
Indicator: A-levels or higher qualification	.1684109	(.1356462)	.0254418	(.1024963)
Indicator: O-levels, or higher qualification	.2260262	(.1244669)*	.257594	(.1028521)**
Socioeconomic status: high (age 33)	.0856948	(.0583498)	.2503771	(.0783291)***
Socioeconomic status: middle (age 33)	-.1489495	(.0902803)*	.0415814	(.0772023)
Parental socioeconomic status at birth: high	.1313775	(.0702612)*	-.0899481	(.0751496)
Parental socioeconomic status at birth: middle	.0068352	(.0611338)	-.0862165	(.0694606)
Years of education: mother	-.0177791	(.0231597)	-.0402418	(.020951)*
Weekly consumption of salads, age 33	.2069069	(.036307)***	.1270041	(.0381515)***
Units of alcohol per week, age 33	-.0112432	(.0215549)	-.0693059	(.0174507)***
Weekly consumption of sweets, age 33	.1037495	(.0316846)***	.0334416	(.0327375)
Dep. Variable: University degree or equivalent (indicator)				
Mathematics test score, age 11 (scores range from 0 to 40)	.0374831	(.0028933)***	.0490307	(.0029136)***
Parental socioeconomic status at birth: high	.3211186	(.0976533)***	.1586365	(.1029511)
Parental socioeconomic status at birth: middle	.2517261	(.0879083)***	-.0197723	(.0875634)
Years of education: father	.0430104	(.0230818)*	.089621	(.0225919)***
Years of education: mother	.1717651	(.0253092)***	.0714152	(.025362)***
Indicator: mother smoker, age 16	-.0325836	(.0321971)	.0121604	(.0337122)
Indicator: father smoker, age 16	-.0798864	(.0293397)***	-.045841	(.0301664)
Indicator: financial hardship, age 11	-.0854498	(.1981986)	-.3247078	(.176241)*
Birthweight	.000183	(.0003946)	-.0010922	(.000487)**
Physical / mental impairments, age 16	-.0373772	(.0200719)*	-.0154767	(.0194875)
_cons	-3,711,185	(.2672855)***	-3,214,851	(.2702207)***
Dep. Variable: A-levels or higher qualification (indicator)				
Mathematics test score, age 11 (scores range from 0 to 40)	.0484951	(.002887)***	.0556811	(.0028521)***
Parental socioeconomic status at birth: high	.3093285	(.096527)***	.365624	(.1028465)***
Parental socioeconomic status at birth: middle	.2016103	(.0832783)**	.0753631	(.0893194)
Years of education: father	.0621672	(.0234857)***	.0978043	(.0218108)***
Years of education: mother	.2105749	(.0271692)***	.0980918	(.0228483)***
Indicator: mother smoker, age 16	-.0638504	(.0326966)*	-.0205379	(.035162)
Indicator: father smoker, age 16	-.0665932	(.028523)**	-.0574186	(.0310675)*
Indicator: financial hardship, age 11	-.0648153	(.1747275)	-.2171577	(.1793203)
Birthweight	.0004401	(.0004534)	-.0001848	(.0005244)
Physical / mental impairments, age 16	-.0432292	(.0196861)**	-.0111454	(.0196192)
_cons	-4,082,782	(.2949768)***	-3,565,177	(.2670044)***
Dep. Variable: O-levels or higher qualification (indicator)				
Mathematics test score, age 11 (scores range from 0 to 40)	.0508223	(.0039558)***	.0388885	(.0031746)***
Parental socioeconomic status at birth: high	.2474022	(.1117653)**	.2832057	(.1020067)***
Parental socioeconomic status at birth: middle	.2336304	(.0840433)***	.0947592	(.0750475)
Years of education: father	.0371301	(.0354716)	.1253908	(.0266867)***
Years of education: mother	.1505105	(.0410317)***	.0309216	(.0374503)
Indicator: mother smoker, age 16	-.1349956	(.0327625)***	-.0416283	(.035538)
Indicator: father smoker, age 16	-.0804349	(.0309384)***	-.1221364	(.0291192)***
Indicator: financial hardship, age 11	-.4688891	(.1256355)***	-.4870313	(.152649)***
Birthweight	.0011249	(.0005735)*	.0004135	(.0006156)
Physical / mental impairments, age 16	-.0029174	(.0204152)	-.0063907	(.0203944)
_cons	-1,340,806	(.4155228)***	-.9199835	(.3991097)**