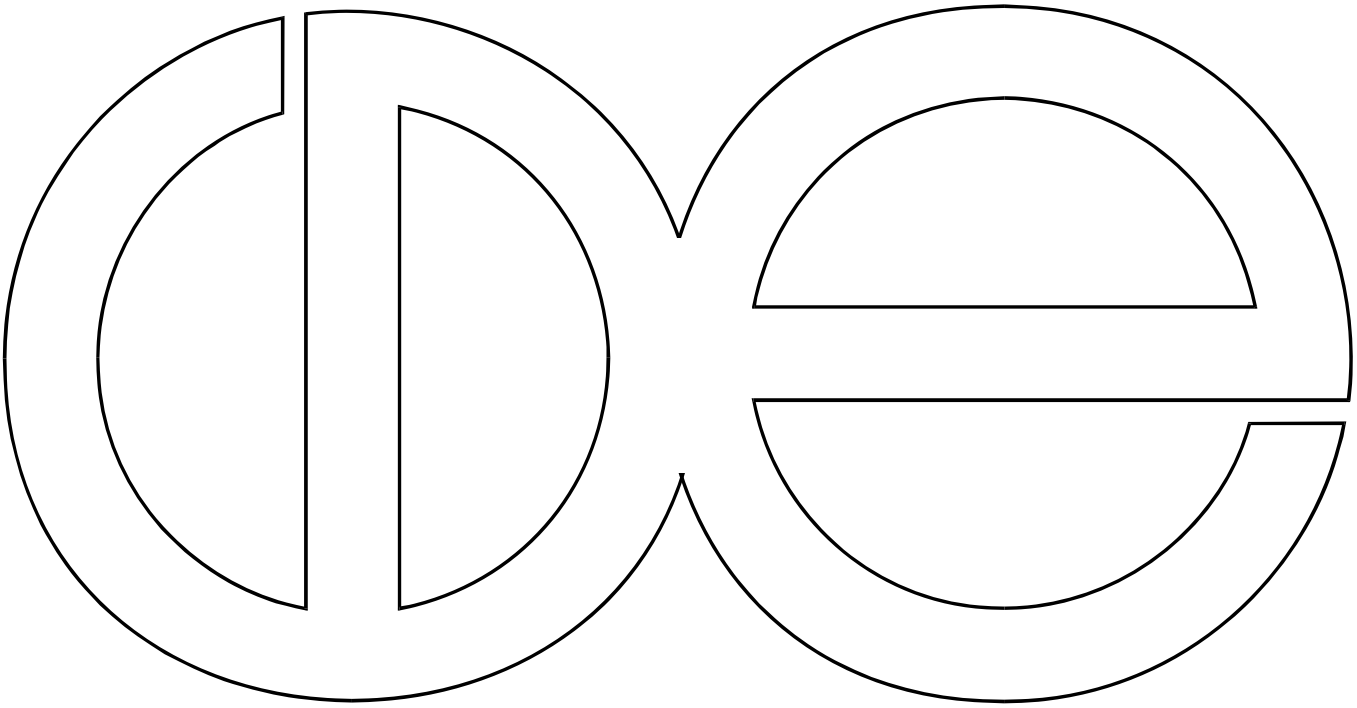
Center for Demography and Ecology

University of Wisconsin-Madison

The Size of Health Selection Effects

Alberto Palloni
Robert G. White
Carolina Milesi

CDE Working Paper No. 2009-10



The Size of Health Selection Effects

Alberto Palloni*

Institute for Policy Research, Department of Sociology, Northwestern University

Robert G. White

Center for Demography and Ecology, Department of Sociology, University of Wisconsin-Madison

Carolina Milesi

Chapin Hall Center for Children, University of Chicago

December 1, 2008

Abstract

The association between adult socioeconomic status and health status has been alternately attributed to the material effects of economic status, ties between labor productivity and health and cumulative effects of early life socioeconomic status. This paper considers the lasting effects of early child health for adult socioeconomic gradients in health status. We analyze longitudinal data with measures of developmental, health and socioeconomic outcomes over the life cycle and show how poor health early in life correlates with lower schooling outcomes, lower adult occupational status and poor adult health. We then examine the cumulative effects of early child health for trajectories of both socioeconomic attainment and health accumulation with a set of counterfactual experiments of health effects over the life cycle. The results show that early child health effects contribute little to the socioeconomic gradient in health that is observed in adulthood.

*Correspondence Address: Institute for Policy Research, Northwestern University, Evanston, IL USA. a-palloni@northwestern.edu.

1 Introduction

The main goal of this paper is to obtain empirical estimates of the contribution of early child health to adult socioeconomic gradients in health. The positive relationship between economic attainment and good health remains a persistent finding across country settings and has proven robust to a wide array of contemporaneous factors. Many models of adult health and of adult economic attainment consistently generate significant estimates of the effects from early child health. However, it is unclear whether these early life effects may also account for an important share of adult socioeconomic gradients in health. We are interested in whether such health gradients reflect selection of individuals into trajectories of socioeconomic attainment according to their early health status. Depending on the cumulative nature of individual health and the dynamics between health and socioeconomic attainment over the life cycle, the existence of child health effects raise the possibility that the origins of adult socioeconomic gradients in health may lie in early life.

Ties between early child health and later economic status have long remained a subject for speculation in explanations of health gradients in the British epidemiological literature (Illsley 1955; Fox, Goldblatt and Adelstein, 1982; Stern, 1983; West, 1991). Early child health is also the main issue of interest in a growing area of epidemiology that examines variants of Barker's (1995) fetal origins hypothesis linking early life to adult chronic disease. Moreover, early child health effects have also been estimated in models of adult socioeconomic outcomes including educational attainment, labor force participation, income and economic status (Power, Matthews and Manor, 1998; Manor et al., 2003; Behrman and Rosenzweig, 2004; Black et al., 2007; Almond, 2006; Palloni, 2006; Palloni et al, 2007; Currie 2007; Mackenbach et al., 1997; van de Mheen et al., 1997; 1998a, 1998b; Case et al, 2002; 2005; Johnson and Schoeni, 2007; Haas, 2006; Haas et al., 2007; Oreopoulus et al., 2008; Duncan et al., 2007). Yet, it remains unclear which mechanisms may link early child health to adult outcomes and the extent to which they might account for observed health gradients in adulthood. This paper evaluates health selection as it may occur through the development of cognitive and non-cognitive skills, the trajectory of individual health status and activities in labor markets. We assess these different paths toward adult health status and socioeconomic status (SES) and the extent to which they may account for observed adult SES health gradients.

The selection effects we examine here concern early life effects on adult status attainment and ought not to be confused with dynamics between health status and economic attainment during adulthood. Dynamics during adulthood involve relatively short-term effects of ill health on wealth (Smith 1999, 2004; Palloni, 2006) and of economic status on access to and

use of health services use as well as the inputs to good health. Nor do we consider here changes in population distributions of health which may be induced by individual frailty (Vaupel et al., 1979). In this case, higher mortality among more frail individuals over time gives rise to contracting, or under rare circumstances, widening adult health gradients. Both of these mechanisms produce associations between SES and health status that do not depend on direct causal effects from health in early life. These are important mechanisms, not to be dismissed lightly, but their manifestation in empirical data does not necessarily reflect the existence of selection effects of the type we examine here.

We study early life health effects on the development of both cognitive and non-cognitive skills and their consequences for educational attainment, adult health and adult SES. We develop a model of individual health and socioeconomic attainment over the life cycle that specifies pathways involving family background, maternal health, individual cognitive and non-cognitive skills and adult outcomes. We estimate the model with the National Child Development Study (NCDS), a cohort of children born in 1958 in Great Britain with measures of child development and family background from birth through age forty-six. We then measure the size of health selection effects over the life cycle with a set of counterfactual experiments of health effects.

Note that significant effects of early health on outcomes such as cognition and educational attainment raise important questions regarding the distribution of opportunities for social mobility as well as the accounting for the full costs of early health. If at all powerful, these relations could establish pathways linking health and economic status across generations and these could play an important role in explaining the stubborn persistence of socioeconomic disparities in health and mortality. Under suitable initial conditions, these pathways could reproduce health and mortality gradients across generations even in open societies with all their baggage of substantial, albeit not perfect, social mobility. Thus, the identification of health selection is substantively important not only for theories explaining the SES adult gradient but also for those concerned with social stratification and intergenerational mobility.

This paper proceeds in section two with a review of theories which tie early conditions to adult health and socioeconomic status. Section three develops a general model of the accumulation of health and socioeconomic attainment over the life cycle that provides a framework for describing the importance of early health for later life outcomes. We describe procedures for addressing measurement concerns in the NCDS and present the model estimates in section four. Section five presents a Monte Carlo method for assessing health selection effects. Estimates of the size of health selection are presented in section six and the final section concludes.

2 Socioeconomic gradients in adult health

2.1 The persistence of health and mortality disparities

There is little doubt that health and mortality gradients by socioeconomic status have persisted and, in some cases, even increased. Much of the evidence adopts educational attainment as a proxy measure of socioeconomic status due to its predictive power for future wages and the paucity of alternative measures of adult socioeconomic status in available datasets. Since the landmark study of Kitagawa and Hauser (1973), studies using different data sources from the United States continue to uncover a recurrent finding–that mortality rates among the better educated are several times lower than those who are less educated (Preston and Taubman 1994; Feldman, Makuc, Kleinman, and Cornoni-Huntley 1989; Lauderdale 2001; Preston and Elo 1995; Elo and Preston 1996; Hummer et al. 1998). Such educational disparities in adult mortality (over age 20) in the United States are tantamount to a displacement of mortality risks by as much as ten years for individuals with lower levels of education (Palloni 2006).

Comparable gradients exist in measures of current health status. For instance, educational gradients in respondents' self-reported health status have been reported in the United States (Palloni 2006; Lynch 2003; Lynch 2006; Smith and Kington 1997a; Smith and Kington 1997b), England and Wales (Banks, Marmot, Oldfield, and Smith 2006), Europe (Mackenbach et al., 2008) and Latin America and the Caribbean (Palloni, McEniry, Wong and Pelaez, 2006). These gradients are strikingly similar in the US, England and Wales, Finland, Hungary, Norway, and Denmark (Valkonen 1987; Valkonen 1989; Valkonen 1992; MacKenbach et al 2008; Sorlie, Rogot, Anderson, Johnson, and Backlund 1992; Rogot, Sorlie, Johnson, and Schmitt 1992. A recent study shows consistent and large income-related gradients in self reported health status across 23 European countries (Eikemo et al., 2008). Finally, a large literature in epidemiology also continues to document socioeconomic gradients in chronic diseases including hypertension, diabetes, coronary heart disease and cancer (see reviews in Mackenbach et al., 1997, 2008).

Socioeconomic gradients in health are also robust to alternate measures of economic status. Gradients tend to become steeper when one uses permanent or even more transient measures of income (Lynch 2003; Pappas, Queen, Hadden, and Fisher 1993; Sorlie, Backlund, and Keller 1995; Lynch 2006; Duleep 1986; McDonough, Duncan, Williams, and House 1997; Duleep 1998). A much stronger relation and steeper gradients are obtained with measures of wealth and assets (Adams et al. 2003; Spittel 2003; Attanasio and Hoynes 2000; Attanasio and Emmerson 2001; Smith 1999; Mare and Palloni 1988; Palloni and Spittel 2004) or even poverty level (Menchik 1993). Using occupational rank-either as prestige scores or with the

standard British classification leads to similar though somewhat more attenuated disparities (MacKenbach et al., 1997; Kitagawa and Hauser 1973; Fox et al. 1982; Mare 1986; Marmot, Ryff, Bumpass, Shipley, and Marks 1997; Moore and Hayward 1990; Power, Manor, and Fox 1991). Furthermore, these differentials remain strong whether one uses mortality measures or self-reported health status and conditions (Palloni 2006; Lynch 2003; Lynch 2006).

2.2 Enduring effects of early childhood health and the possibility of health selection

While the empirical record provides consistent evidence of health and mortality gradients that are large, persistent and somewhat insensitive to alternative metrics of SES, there is less of a consensus over the mechanisms which may underlie these relationships. Is income, wealth or education health-protective? Are higher social strata or classes endowed with resources or traits that reduce exposure to risks, increase recovery rates and reduce lethality of illnesses? If so, which resources are relevant and what are the mechanisms linking SES to health?

We are interested in a health selection mechanism in which health status early in life selects individuals during childhood for different trajectories of SES attainment. Suppose individuals are allocated to various health statuses at birth via some mechanism or even at random. Suppose further that individuals at the lower end of the health distribution are also less likely to perform well in elementary school, more likely to have a poor track record in high school, less likely to participate in social activities and, more generally, less able to acquire or retain traits that are both relevant for socioeconomic attainment and conducive to a healthier lifestyle. It may well be that lower SES by itself increases the probability of adopting unhealthy behaviors (such as drinking, smoking, lack of exercise and neglect of preventative care). However, such effects occur in addition to the fact that an individual with relatively lower socioeconomic attainment is also more likely to have been drawn from the lower end of the health status distribution. Not only are these individuals more likely to develop conditions that make them susceptible to even worse health in adulthood but weaker health itself becomes a constraint in the quest for upward mobility. The process may proceed through feedback effects and cumulative damage and disadvantages whereby lack of health begets fewer resources and diminished resources begets unhealthy states. While these relationships may take a lifetime to unfold and they may be difficult to identify empirically in cross-sectional or conventional panel studies, they could well underpin observed contemporaneous associations between SES, health and mortality over the life course.

Three remarks are in order. First, the effective operation of health selection does not

require that early health status be allocated via a specific mechanism. It can operate even if early health is allocated randomly. However, the *intergenerational transmission of health status and SES* via health status can only take place if early health status is allocated non-randomly. Were health selection to be pervasive, it will sustain the self-reproduction of health gradients across generations if and only if antecedent health status is distributed non-randomly.

Second, while the existence of mechanisms whereby early conditions directly influence adult health could reinforce selection, they are neither necessary nor sufficient for health selection to operate efficaciously. If, for example, the conjecture about a critical period is borne out by the facts and some events experienced early on in life do indeed lead to higher risks of congestive heart disease or diabetes II (Barker 1992, 1995), this, by itself, does not automatically result in health selection. By the same token, health selection does not require the existence of critical period effects to be an effective producer of SES gradients. The necessary ingredient is that, at one or more points in the life course, acquisition of the traits which are required to achieve relatively high SES is a function of antecedent health status.

Third, if health selection does indeed take place, it involves both effects of health on socioeconomic status and of socioeconomic status on health. Selection neither excludes nor is inconsistent with the idea that a regime of socioeconomic strata can manufacture health and mortality disparities. Quite the contrary: it represents a mechanism through which such disparities and the social stratification system itself may be reproduced over time (Palloni et al., 2007).

The possibility that a health selection mechanism could contribute to the ordering of individuals into adult socioeconomic and health status was prominently noted in the Black report (Black et al., 1988) where it was described as a process of indirect selection.[1] Such a process has long been proposed in the social epidemiology literature (e.g., Illsley 1955, 1986; Stern 1983; West 1991) where it first emerged as a point of speculative hypothesizing but has since become a subject for increasingly direct empirical tests (Power, Manor, Fox, and Fogelman 1990; Power, Manor and Fox 1991; Vagero and Illsley 1995; Power and Matthews 1997; Power, Hertzman, Matthews, and Manor 1997; Blane 1999; Blane, Davey Smith, and Hart 1999; Martorell, Stein, and Schroeder 2001; Chandola et al., 2003; Manor, Matthews and Power 2003; Kuh and Ben-Shlomo 2004, Palloni and Ewbank 2004). Empirical studies

[1] Given our understanding of this process, it is odd to refer to it as emphindirect selection or emphselection. It is one among many mediating pathways that involves effects and feedbacks where health status and SES alternate as constraints on the outcomes. Even though unsatisfying, in the remaining of this paper we preserve the terminology, however ambiguous, and continue to use the label emphselection effects solely to avoid confusion.

in this area largely conclude that health selection mechanisms are of rather muted importance, and that the bulk of the observed health gradients must be accounted for by either attributes of socioeconomic strata themselves or by the effects of health status on wealth accumulation.[2] However, in the absence of an appropriate theoretical model linking early health status with adult socioeconomic attainment, these studies fail to offer fully specified tests for distinguishing health selection from other mechanisms that explain contemporaneous correlations between adult SES attainment and health status. The case is not yet closed.

Recent research in economics contributes with evidence of the long lasting effects of child health on adult socioeconomic outcomes. For instance, evidence of the long-lasting effects of the 1918 influenza pandemic offers a novel test of a variant of Barker's fetal origins hypothesis (Barker, 1992) and shows striking correspondence between in utero exposure and adult educational attainment, wages and disability status (Almond, 2006). Twin studies which take explicit account of the unobserved effects from both family background and genetic endowments report significant birth weight effects on later cognition and labor market outcomes (Behrman and Rosenzweig, 2004; Black, Devereux and Salvanes, 2007; Oreopoulos et al., 2008). However, these studies are limited to measures from early childhood and adulthood only, a long time period during which we expect much of the process of health selection may unfold. The estimates from twin samples make important contributions by establishing both the considerable magnitude of unobserved family background effects as well as the robustness of early child health effects to such unobserved dimensions. Yet, these estimates must be qualified by their small sample sizes (e.g., Behrman and Rosenzweig, 2004; Oreopoulos et al., 2008) and the unexplained instability in estimates across cohorts (e.g. Black et al., 2007). As readily acknowledged in this work, the lower average birth weight of twins also calls into question how representative these samples may be of the general population of singleton births. Most importantly, without the benefit of measures of adult health status, these studies adopt body mass index, a coarse approximation of health at best which allows little insight into the adult health gradients which are at the center of our study.

Estimates of early child health effects on adult outcomes with the NCDS birth cohort have made particularly important contributions to understanding the lasting effects of child health. The origins of adult gradients in this cohort have been shown to reach back to early childhood (Case, Lubotsky and Paxson, 2002) and estimates of child health effects have

[2] For the most part reverse causality has been shown to be non-negligible only at relatively older ages (Smith, 1999) and even here there are lingering doubts about how strong the reverse effect of health on SES is (Adams et al., 2003)

further withstood adjustments for early family economic status and intervening adolescent health (Case, Fertig and Paxon, 2005). Our work examines related questions, but departs from these studies in three ways. First, we establish a set of hypotheses that evaluate specific mechanisms linking early health to educational attainment and labor market outcomes. We emphasize the development of cognitive and non-cognitive skills for educational attainment and employment. Second, we address the sample selection problems in the NCDS which we believe bias estimates of early child health on later life outcomes. Third, we undertake a set of counterfactual experiments of health effects to assess cumulative health selection over the life cycle. These steps allow us to evaluate one set of mechanisms which may underlie health selection and the extent to which they may account for adult health gradients.

3 A Model of Health and Adult Attainment

We propose a model that considers early life health effects on the development of both cognitive and non-cognitive skills and their consequences for educational attainment, health status and adult socioeconomic status. Significant effects from early health on such outcomes are relevant for understanding social mobility and the mechanisms that may link health and economic status across generations. Selection effects require the existence of causal pathways between early conditions and health status, on the one hand, and adult socioeconomic status and health, on the other. The model we propose identifies some of these pathways. We develop a stylized model that extends Grossman's (1972) original formulation of health production. We adapt Cunha and Heckman's (2008) framework for human capabilities development and emphasize cognitive and non-cognitive skills in the accumulation of health and socioeconomic attainment. Our model builds upon Todd and Wolpin's (2007) general formulation of cumulative attainment and remains agnostic about the precise processes underlying skills development.

We begin with a general production function of status attainment (A) for individual i at age a. Status attainment depends on a vector of prior inputs into individual human capital $K_i(a)$ that may occur at any time up to age a and an inherited endowment of attributes p_{i0}:

$$A_{ia} = A_a(K_i(a), p_{i0}) \tag{1}$$

In this case, $K_i(a)$ includes both current and preceding health status as well as exogenous effects on individual status attainment.

An approximation in which status attainment is linear in a vector of inputs X_{ia} which

do not depend on child age follows:

$$A_{ia} = X_{ia}\alpha_1 + X_{ia-1}\alpha_2 + ... + X_{i1}\alpha_a + \beta_a p_{i0} + \nu_{ia}\rho_1 + \nu_{ia-1}\rho_2 + ... + \nu_{i1}\rho_a + \varepsilon_{ia} \quad (2)$$

Current status in this case is also a function of omitted current and lagged inputs ν_{ia} and measurement error. A common approximation of this model is the following:

$$A_{ia} = X_{ia}\alpha + \gamma A_{ia-1} + \epsilon_{ia} \quad (3)$$

In this case, the implied assumptions concerning the returns to lagged and current inputs distinguish the underlying technology for attainment. [3]. The nature of the NCDS data and the lack of a plausible instruments prevent us from addressing the effects of unobserved variables, changes in unobserved effects over time and the possible dependencies between such effects and inputs. However, adding lagged inputs captures the changes in the effects of inputs over time without any assumption about the rates of change.

By further adding proxy measures for individual endowments p_{i0}, the following less restrictive model of status attainment results:

$$A_{ia} = X_{ia}\theta_1 + X_{ia-1}\theta_2 + \ldots + X_{i1}\theta_a + \gamma A_{ia-1} + \beta_a p_{i0} + \epsilon_{ia} \quad (4)$$

To allow separate processes producing economic status and health status, equation (1) above may be expanded with a vector A_{ia} that includes both socioeconomic status and health status in a formulation similar to Cunha and Heckman (2008). In this case, inputs are distinguished by their effects in acquiring socioeconomic status and health status:

$$\begin{pmatrix} S_{ia} \\ H_{ia} \end{pmatrix} = A_a \begin{pmatrix} K_i^{se}(a), s_{i0} \\ K_i^{h}(a), h_{i0} \end{pmatrix}$$

Here, the set of endowments p_{i0} includes traits that affect socioeconomic attainment s_{i0} and those which affect health status attainment h_{i0}. Equation (4) then may be expressed as two separate equations generating socioeconomic and health status.

$$S_{ia} = E_{ia}\lambda_1 + E_{ia-1}\lambda_2 + \ldots + E_{i1}\lambda_a + \gamma^{se} S_{ia-1} + \beta_a^s s_{i0} + e_{ia}^s \quad (5)$$

$$H_{ia} = I_{ia}l_1 + I_{ia-1}l_2 + \ldots + I_{i1}l_a + \gamma^h H_{ia-1} + \beta_a^h h_{i0} + e_{ia}^h \quad (6)$$

To specify selection effects over a lag of length j, we may include health status in the appropriate vector of lagged inputs E_{a-j} into socioeconomic status. Likewise, we may include

[3] The assumptions are described in detail in Todd and Wolpin (2007)

socioeconomic status as a lagged input among the inputs I_{a-k} into health status over any lag of length k. Incorporating these dynamic effects in both models of attainment then gives rise to our main models of interest.

$$S_{ia} = E_{ia}\lambda_1 + E_{ia-1}\lambda_2 + \ldots + E_{i1}\lambda_a + \gamma^{se,se}S_{ia-1} + \gamma^{se,h}H_{ia-1} + \beta^s_a s_{i0} + e^s_{ia} \tag{7}$$

$$H_{ia} = I_{ia}l_1 + I_{ia-1}l_2 + \ldots + I_{i1}l_a + \gamma^{h,h}H_{ia-1} + \gamma^{h,se}S_{ia-1} + \beta^h_a h_{i0} + e^h_{ia} \tag{8}$$

Incorporating prior status attainment as a lagged input implicitly accounts for dynamic effects between health status and socioeconomic status over the life course. This framework also takes account of the limitations of survey data in which inputs may be only periodically collected and defined by criteria that vary across waves. Moreover, this framework is also suited to account for dynamic effects between socioeconomic and health status by allowing correlated errors across the two models.

In this paper we adopt family SES $S_{i,0}$ and mother's health at birth as proxy measures for inherited traits s_0 and h_0. Attainment of SES at age a (S_a) and of health at age a (H_a) depend on S_{a-k}, H_{a-k}, prior inputs, endowments at birth, and shocks experienced in the age interval $(a-k, a)$. We further approximate equations 7 and 8 given that the NCDS offers limited measures of conditions, inputs and attainment at only selected stages in the life cycle of individuals. The availability of adult waves in the NCDS also limits the number of possible lags of adult attainment to two (ages 33 and 41). Adopting A^z_a (where $z \in \{S, H\}$) to represent attainment of either S_a or H_a during the three latest waves of the NCDS (ages 33, 41 and 46), the production of adult attainment with the available measures in the NCDS may be expressed by the following:.

$$\begin{aligned} A^z_{i,a} \;=\; & \lambda_0 + \lambda_{11}\, Cognitive\, Skills_{i,11} + \lambda_{12}\, Non\, Cognitive\, Skills_{i,11} + \gamma^{z,z}_1 A^z_{i,a-k} \\ & + \gamma^{z,h}_1\, Chronic\, Conditions_{i,16} + \gamma^{z,h}_2\, Chronic\, Conditions_{i,7} \\ & + \gamma^{z,h}_3\, Birth\, Weight_{i,0} + \gamma^{z,se}_1 Education_{i,23} + \beta^s_a S_{i,0} + e^s_{i,a} \end{aligned} \tag{9}$$

The diagram in Figure 1 makes explicit the relations between the main inputs in the model. This model captures the most important relations that could result in health selection.[4] The model that we estimate is a mixture model where we include lagged values of $S_{i,a}$ and $H_{i,a}$ on each other as well as explicit effects of selected health-relevant inputs on capability relevant inputs. Although not shown in Figure 1, additional controls including current

[4] Elsewhere we have provided a full rational for the model which involves elements from labor economics, social stratification, child development and recent advances in epidemiology (Palloni et al., 2007).

SES over adolescence are also included. It is a compromise model that remains consistent with the abstract model in (1) but conforms to limitations imposed by the data set we use. In particular, we distinguish five stages in the life cycle of individuals: birth, early childhood, adolescence and early and late adulthood. As shown in Figure 1, health at birth, H_0, is an outcome of both family background S_0 and mother's health at birth (a proxy measure for an endowment of health).

Our main equations of interest to examine the pathways to adult health gradients concern the development of cognitive and non-cognitive skills. We adopt general formulations for the development of cognitive and non-cognitive skills (inputs E_a) which depend on family background, mother's health, early child health (H_0 and H_7) and adolescent health (H_{16}). A general equation for $Skills^y_{i,11}$ development (where $y \in \{Cognitive\,Skills\,,\ Non\,Cognitive\,Skills\}$) encompasses both equations:

$$\begin{aligned} Skills^y_{i,11} \;=\; & b_{y0} + b_{y1} S_{i,11} + b_{y2}\, Parents'\, Education_{i,11} + b_{y3}\, Attractiveness_{i,11} \\ & + b_{y4}\, Chronic\, Conditions_{i,7} + b_{y5} S_{i,7} + b_{y6} Birth\, Weight_{i,0} \\ & + b_{y7} S_{i,0} + b_{y8}\, Maternal\, Health_{i,0} + \epsilon^c_{i,11} \end{aligned} \tag{10}$$

These conditions are, in turn, the main inputs that determine educational attainment, occupational status and health status at ages 33, 41 and 46. We also specify early childhood and adolescent health at age 7 and age 16 respectively (chronic conditions) as functions of health at birth and family background. In addition to models 8 and 9 characterizing cognitive and non-cognitive skills at age 11, there are then eight additional models that describe health and SES attainment over the stages of the life course for which we have information. [5]

Our primary focus of analysis is the relation between adult SES and health status measured at ages 33, 41 and 46. Although the model represented by the path diagram is constrained by data availability, it reproduces key pathways linking health and socioeconomic status at points in the life cycles for which we have available information. In particular, we may adapt the model to estimate relationships between early life events and later life outcomes for any stage of the life course which may be observed. The limitations presented by the reach of our data over the life course are clear: since most health events occur after age 45 or 50 we are likely to underestimate the total adult health gradient. However, limiting the analysis to this age range also has an important advantage: by truncating the life course at age 46 we minimize problems engendered by reverse causality and unmeasured

[5] Throughout we assume that constructs are well measured by a single indicator and we do not at this time consider latent constructs.

heterogeneity, the effects of which play out more strongly later on in life.

4 Health and Economic Attainment over the Life Cycle

4.1 Data

The National Child Development Study (NCDS) is a prospective longitudinal study of nearly all (98 percent) children born in the week of March 3-9, 1958 in Great Britain (England, Scotland, and Wales). Medical, social, demographic and economic data have been collected on cohort members from birth through age 46, including follow-up interviews their parents, teachers, doctors, or their partners at ages 7, 11, 16, 23, 33, 41 and 46.[6]. As illustrated in Figure 2, the NCDS cohort has displayed an increasing health gradient over time that is consistent with many of the reports discussed above. Figure 2 illustrates one dimension of the health gradient by reporting the distribution of individuals in self-reported poor health across the six categories of self-reported SES in the NCDS. These gradients are calculated for the observed values of socioeconomic and health status at ages 33, 41 and 46. The measure of SES is a six point scale that is constant across all waves of the survey. Respondents' self-reported health assessments which were either fair or poor define the indicator of bad health. The negative slope in the relationship between current socioeconomic status and health is evident at all ages for which there are available measures of health and socioeconomic status in the NCDS. The weak relationship at age 33 and increasing slope in the relationship over ages 41 and 46 is consistent with the emergence of the health gradient in middle adulthood.

As in most longitudinal surveys of comparable duration, survey non-response is a serious concern in the NCDS. While it has been reported elsewhere that sample refreshment sufficiently maintained the representativeness of the NCDS across all waves (Plewis et al., 2004; Hawkes and Plewis, 2006), we find that survey non-response leaves a select sample by wave six with respect to the measures that we consider in our estimates of early life health effects. Our sample of interest is all 8,998 boys present at baseline (48.5 percent of the full sample). We limit the sample to boys to avoid the complications of accounting for the differences in labor market participation between men and women during middle adulthood. As may be expected, the initial loss of 5.3 percent of this sample due to child deaths is predicted by birth outcomes such as birth weight, mother's smoking status and family socioeconomic characteristics. To the extent then that low birth weight affects later life outcomes, we expect that this sample selection biases our estimates of early health measures downward. The magnitude of survey non-response throughout the duration of the NCDS is large. Of our

[6] For a detailed description of the study see Ferri (1993) and Blane (1999)

remaining sample of 8,512, while only five percent of this sample permanently attrited before wave three, 61 percent failed to participate in at least one wave. Given that our model includes measures from all available waves, this magnitude of survey non-response incurs serious consequences for our sample–combined with item non-response, the case complete sample for estimating the model outlined in Figure 1 diminishes to 1,987.

By any standards a reduction of the original sample of such magnitude should constitute a call for caution. While concern over the resulting scope for sample selection bias may be found in related studies using the NCDS, case complete samples are often adopted under the assumption that refreshment samples adequately compensate any resulting changes in the sample attributes of concern.[7] As shown elsewhere (White, 2008), the majority of the reduction in our sample size is due to survey non-response which may be estimated with an additional set of variables from the NCDS. A series of logistic models predicting wave non-response are estimated with a set of variables that are auxiliary to our model variables. The large degree of variation in survey non-response explained by these models suggests that some of the effects of sample selection may be then be corrected with these sets of auxiliary variables.

For these reasons, rather than estimating parameters on the subsample of case complete respondents, we multiply impute missingness in our variables due to either survey non-response or to item non-response.[8] We adapt a Bayesian imputation approach originally proposed by Rubin (1976; 1987). These procedures require that missingness is ignorable: the pattern of missingness for each variable does not depend on the value of the missing item after conditioning on other observed variables available in the data; and, any parameters which may govern missingness may not be related to the parameters in our models of interest. To the extent that different types of missingness are predicted with the available data in the NCDS (e.g., permanent attrition, refusals, temporal attrition, item missingness) our estimations for multiply imputing missingness in the NCDS take into account many of the observed factors that explain missingness. Consequently, we maximize the opportunities for satisfying the assumption of ignorable missingness. We impute categorical and continuous variables, implementing normalizing transformations for several variables where necessary. We used a large set of auxiliary variables in the imputation of missing values which were selected by a model search routine including multivariate analyses of non-response probabil-

[7] Closer attention has been devoted to handling item non-response, particularly in the available measures of family income (Case et al., 2005). However, the dummy variable adjustment for missingness that is adopted in these cases remains only a partial adjustment and raises further opportunities for bias even when missingness occurs completely at random.

[8] For details on the models used for imputation and a full analysis of the nature of attrition in the panel, see White (2008). The multiple imputation models used in this paper take into account the nature of attrition uncovered by these analyses.

ities and a broad sweep over all the variables available in the NCDS. A total of ten imputed data sets were determined to sufficiently capture the between-sample variance in our model estimates across multiple imputed datasets.[9]. To ensure that we could impute values for our main covariates of interest we include multiple imputed values for missingness in all variables. But we do not include imputed values for health and socioeconomic status at age 46 in the sample used for the analysis.

4.2 Measuring early childhood health and SES attainment

We adopt an indicator of low birth weight as the earliest measure of child health. There is ample reason to be concerned that the contrast low/normal birth weight is a coarse indicator of health status at birth. Birth weight is determined by both intrauterine growth and gestation length, distinct clinical concepts with different underlying causes which have been associated with different health consequences in later life.[10] There may be concern that this general measure of early health risks mixing differential effects of different underlying causes. However, we found that the main results and key inferences are not altered regardless of whether or not we use length of gestation in combination with a continuous measure of birth weight or a low birth weight indicator to measure early health. For this reason we settled for using birth weight as the main indicator for very early health status.

There are other reasons why this choice could be problematic, however. In particular, in the absence of a random allocation of investments in early child health, it is difficult to infer the degree to which unobservable investments in child health which correlate with socioeconomic characteristics as well as investments in other dimensions of human capital may bias our estimates of the effects of health on later life outcomes. Evidence from twins studies of the large share of the effect of birth weight that is attributable to unobservable family background characteristics and genetic endowments (Almond et al., 2005) suggests that such biases may be formidable. However, these studies also confirm that birth weight continues to exert a significant and large effect on later life outcomes even after adjusting for unobservable family characteristics (see also Currie, 2007). Moreover, we also consider measures of chronic conditions at two later periods (ages 7 and 16) that should provide additional indicators of true latent health status. While estimates of these measures are arguably subject to the same source of bias as estimates of birth weight effects, both their direction and relationship to the bias in birth weight estimates are unclear. For these reasons,

[9] We use a sequence of estimations for imputing sets of missing values (Raghunathan et al., 2001) and adapt a data augmentation routine (Tanner and Wong, 1987) to generate multiple datasets including imputed values

[10] See Huxley et al. (2002) and Victora et al. (2008) for reviews of the different later life consequences of different measures of birth weight.

we adopt birth weight and measures of chronic conditions during adolescence as indicators of underlying health, but temper our interpretations of the reduced effects of birth weight on adult outcomes.

Finally, the theoretical argument about the influence of child health is quite general and does not refer to physical but also to psychological and emotional help. In this paper we make use of birth weight and parents' reports of chronic conditions at two points in time during adolescence. The number of chronic conditions is the sum of conditions including diabetes, asthma and conditions related to neurological health, cardiovascular health, hearing and vision. Admittedly these indicators are a very blunt tool to capture the complexity of the early child health dimension and we cannot hope to fully capture the complex phenomenon we are trying to study. However, the availability of such periodic measures of health status over adolescence offers a general indication of the early trajectory of health status.

The NCDS measure of economic attainment is a self-reported SES assessment along a six-point scale according to occupational ranking. This classification was originally devised by the British Registrar General for use in the national census and has been widely used in many areas of research, particularly among those working on health and mortality disparities. [11]. The NCDS also includes measures of wages at various ages and these too could have been used to construct an indicator of SES. However, we chose to focus on the SES scale for two reasons. First, the available wage data in the NCDS were judged to be inadequate for our purposes since it was not collected from parents at the baseline and later measures of household income and cohort member wages contain considerable item missingness. Patterns of this missingness vary both within and across waves and missingness is highly correlated with individual health and socioeconomic measures, raising concerns about selection bias in models of health including wages. Second, many of the epidemiological studies of heath selection reviewed in section two as well research in economics that uses the NCDS to study adult economic outcomes (e.g. Case et al., 2002; 2005) commonly employ SES as a metric for economic attainment. Had we chosen wages as our preferred indicator of SES we would have precluded comparisons of our results with those from extant studies. [12]

[11] Brewer (1986) details the history of the Registrar's SES measure and its uses in social science. See Fitzpatrick and Dollamore (1999) and Drever and Whitehead (1997) for reviews of studies in health that adopt this measure of SES

[12] We experimented with re-estimating all models using wages as the indicator for SES. By and large, we find that these models fit much worse that those using the six-point scale SES and that the behavior of key determinants is erratic. We attribute this to the severe degree of missingness and likely high level of measurement error.

4.3 Model estimates

The path diagram in Figure 1 translates into a standard set of non-recursive structural equations. Each of the constructs in the path diagram is represented by a single indicator drawn from the data set (a full description of these appears in Table 1). The parameters of the system can be estimated via maximum likelihood procedures to enable us to handle equations where the dependent variables are discrete. Because of the use of multiple imputation, we must estimate the corresponding structural equation model in each one of ten alternative imputed data sets. These results are then averaged and each average estimate thus obtained is associated with standard errors which are corrected to account for both the within-sample and between-sample variance in the estimate. Reported model statistics are similarly averaged across the multiple datasets.

Table 2A reports estimates for six models of intermediate outcomes between birth and early adulthood. Birth weight is estimated as a logistic function of family socioeconomic status at birth and measures of maternal health and prenatal behaviors. Chronic conditions at ages 7 and 16 are estimated as negative binomial regressions of SES at birth, current SES, and birth weight. Chronic condition at age 16 also includes the lagged count of chronic conditions. The measure of cognition is a standardized mean of a set of math and reading exams; behavior is a sum of teacher assessments of students' behavior along five different dimensions. Both outcomes are estimated as ordinary least squares regressions and include birth weight and chronic conditions as independent variables. Education is measured by a five point scale derived from the numbers of passed O-level and A-level exams, completed higher education and professional certification and is estimated as an ordered logit. [13]

The estimates from these six models illustrate the paths by which early health effects may shape trajectories for later socioeconomic attainment. Confirming results from other research, an indicator of low birth weight is significantly associated with lower cognitive scores. Chronic conditions at age 7 is significantly associated with the natural log of the non-cognitive skills measure and also an important predictor of chronic conditions at age 16. The large significant effects of both cognitive and non-cognitive skills in the model for education illustrate the pathways that tie birth weight and chronic conditions at age seven to educational attainment. While birth weight and chronic conditions do not exert additional direct effects on education, their effects through these skills pathways begin to illustrate the

[13] Model statistics represent the mean values of the relevant statistic across the ten imputed datasets. Minimum and maximum values are also included to indicate the variance in model fit. The standard rules for correcting estimates from multiply imputed datasets to account for within and between sample variance (Rubin 1987) rely on asymptotic normality. Consequently, the geometric mean is adopted for the R-square estimates since the sampling distributions of the log R-square is closer to normality than the sampling distribution of R-square. Arithmetic means are adopted for the remaining statistics.

mechanisms linking early health to adult attainment.

Cognitive and non-cognitive skills also play important roles in the cumulative effects of family SES. The significance and magnitude of the effects of SES over childhood are consistent with cumulative effects of family socioeconomic background. SES at birth is significant in the models for chronic conditions at age 7, cognitive skills and education. The robustness of the effects of this early measure of family background to current family SES in the models for cognitive skills and educational attainment is consistent with cumulative effects from occupancy in low SES groups. Moreover, the significance of both lags of SES independent of cognitive skills in the education equation illustrate further accumulation of the effects from early family background.[14] Non-cognitive skills present an additional pathway for cumulative effects of family SES at age 11 in educational attainment.

Table 2B reports estimates for the models of socioeconomic status and health outcomes at ages 33, 41, and 46. Adult health and socioeconomic status are estimated as functions of all prior outcomes including SES at birth and, in the case of outcomes at ages 41 and 46, lagged values of both SES and health. SES is estimated with an ordered logit while an indicator of self-reported poor health is estimated as a logit.

The models for SES at ages 33, 41 and 46 illustrate the importance of skills development in the reach of early health to adult economic attainment. Early health displays few direct effects on adult SES. Neither low birth weight nor chronic conditions at age 16 are significant for any of the adult measures of SES. While there is a significant effect of chronic conditions at age 7 on socioeconomic status at age 41 (and a weak effect at age 46), these effects are ambiguous given the absence of additional effects on either prior or later SES as well as the lack of increasing effects with the number of chronic conditions at any ages. However, the large effects of cognition and education illustrate the pathways by which early health may indirectly affect adult SES. Both sets of effects are large in all three adult SES models and persist with controls for SES at birth and lagged SES. The large effects of education also increase across the indicators of increasing levels of completed education. The significance of cognition in all three adult SES models in addition to its significance in the education model underscores how the early health effects on cognition shown in Table 2A may have cumulative effects in selecting individuals for different trajectories of socioeconomic attainment. The absence of direct effects of non-cognitive skills in these models suggests that their importance in the pathways of health effects may be limited to education.

The models for health status at ages 33, 41 and 46 illustrate the importance of adult

[14] The measure of current SES in the education equation is family SES at age 16. This may be interpreted as lag SES given that the measure of educational attainment includes measures spanning adolescence (O-level and A-level exams) to adulthood (university education).

health status in additional pathways linking early health to both health and socioeconomic attainment in adulthood. As in the equations for adult SES, low birth weight and the two chronic conditions measures do not display significant associations with adult health. The absence of such effects suggests that any trajectory of health accumulation over the periods recorded in the NCDS is not consistent with the trajectories described in the fetal origins hypothesis (Barker, 1992). Rather than direct health effects, education presents an important tie between early and adult health. Confirming other research on the subject, education displays significantly negative associations with poor health at all three ages. These effects are also largely increasing in magnitude with gains in education. The education gradient tends to attenuate with age but even at age 46 there are still some important protective effects for those with higher levels of education. Remarkably, these effects are robust to the inclusion of controls for lagged SES.

Lagged SES presents another important pathway linking early health effects to adult health status. We confirm the importance of the lagged SES gradient of health at age 46.[15] For individuals in SES group 2, the odds of reporting poor health are 1.66 (exp(.504)) times as large as for individuals in the highest SES group. The effect for individuals in SES group 4 is lower, although still substantially larger than for those in the highest SES group (odds = 1.52).[16] To the extent that early health imparts selection effects for adult socioeconomic attainment, these effects may further feedback to poor individual health in adulthood.

The significance of both education and adult SES for adult health illustrate further dimensions by which early health may affect adult health. These results also suggest that the most salient and persistent health disparities are those across levels of education rather than SES. However, these effects stand to impart the largest consequences for the health gradient through dynamics between health and SES *during* adulthood. The coefficients for lagged health are highly significant for SES at age 46 as well as for current health at both ages 41 and 46. The significance of both lags of health for health at age 46 further raises the possibility of cumulative damage from adverse health in adulthood in addition to the likely effects of unobserved heterogeneity. Such dynamics may be especially important in the ties between early health and the gradients in adult health defined by both education

[15] The lack of significance of the lagged gradient with the full set of controls presented in Table 2B is attributed to the compressed distributions of both SES at age 33 and of health at age 41. Earnings trajectories in mid-life that accelerate in the late 30s likely underlie the increase in the spread in the SES distribution by age 41.

[16] An even larger gradient may be observed in health at age 46 with respect to SES at birth.

and SES.[17].

The models in Table 2B depart from conventional studies of health and economic attainment in the literature by including lagged health status and emphasizing the lagged SES gradient in health rather than the contemporaneous gradient. In order to ensure that our results are comparable to prior studies, we estimate additional models of the contemporaneous gradient that are comparable to those available in the standard literature. We estimate alternative models for health status at ages 41 and 46 with lagged SES where we alternatively control or ignore lagged health status. Similarly, we estimate models including contemporaneous instead of lagged SES and no controls for lagged health status. The results are displayed in Tables 3A-3B.

The left panel of Table 3A illustrates modest attenuation in the lagged SES gradient in health at age 46 when lagged measures of health are included. While including one or two lags both shifts the gradient and alters its slope, the effects demonstrate the robustness of the lagged SES gradient in health at age 46 to lagged measures of health. As shown in Table 2B, our model for health at age 41 does not demonstrate a significant gradient with respect to lagged SES. However, the education gradient is similarly robust to the lagged measure of health. The left panel in Table 3B reports both the education and SES gradients for health at age 41. While there is greater attenuation in the education gradient for health at age 41 than in the SES gradient for health at age 46, the effects remain substantial and increasing with the level of education. The effects on the conventional contemporaneous gradient are more severe. We confirm the large effects that are common in reports of contemporaneous SES gradients. The right panel of Table 3A shows that membership in the lowest social class increases the odds of falling into poor health at age 46 by about 1.85 (exp(.618)) relative to membership in the highest class; the right panel of Table 3B shows that at age 41 the increase is 2.15 (exp(.766)). This reflects a sharp gradient that is robust to the traits that are associated with both socioeconomic and health attainment. It also illustrates how much of the contemporaneous SES gradient in health may also reflect the reverse effects of health on current SES. In Table 3A, including one or two lags of health in the model of health at age 46 attenuates the SES gradient by as much as 28 percent (SES group 1). While

[17] All models in Tables 2A and 2B were also estimated with the case complete data to examine the sensitivity of the results to the adopted method for handling missing data. The expected loss in precision in the estimates with the case complete data does not alter any of the results reported above. There is also no clear pattern of changes in the magnitudes of estimates across the full set of covariates from all 12 models. However, increases in the magnitudes of key estimates in the analysis illustrate the possible scope of sample selection bias in estimates with the case complete data: family SES in the model of non-cognitive skills; cognitive skills in the education model; and, both cognitive skills and education for SES at age 33. These larger effects are consistent with the effects of sample selection along the key inputs for adult SES attainment. A full set of sensitivity analyses of the sample selection correction method adopted in this paper is outlined elsewhere (White, 2008).

these results demonstrate the robustness of the lagged SES gradient and education gradients to lagged health, they also caution interpretations of SES gradients without the benefit of lagged measures of health.

In summary, these results suggest three inferences. First, the significant effect of birth weight on cognition suggests a health selection effect that occurs along a pathway dominated by cognition, education and the dynamics of health and SES in adulthood. Second, early health measures do not display direct effects for adult health and socioeconomic status independent of their effects on preceding outcomes in the life course. Third, the education and SES gradients in adult health remains substantial even after accounting for early child health traits and other traits which may depend on early childhood health. Are these results adequate to support an inference centered on health selection namely, that the effect of birth weight on cognition represents a possible path along cognition and education by which early health may influence adult socioeconomic and health status?

This conjecture cannot be tested with the above results for three reasons. First, the estimates reported above (and those from any structural equation model) only reflect the strength of associations between pairs of conditions (given the model). Model estimates for outcomes at different stages of the life course do not address the uncertainty about the effective magnitude of the contribution of selection effects to the total socioeconomic gradient in health. The estimates presented in Tables 2A and 2B reflect only piecewise direct estimates of associations of selected covariates of interest with the chosen outcomes. Second, the estimates are insufficient to gauge the magnitude of cumulative effects associated with complicated pathways.[18] Third, the estimates cannot account for changes in the distribution of individuals who experience those pathways throughout their life cycle. The estimates ignore the initial and subsequent distributions of traits across individuals and over the life course as characterized by the path model. For these three reasons, our model estimates by themselves cannot reveal the effective magnitude of the contribution of selection effects to the total SES adult health gradient.

5 Measuring health selection

We propose a simulation method that assesses the magnitude of health selection effects on adult socioeconomic and health status. We implement a procedure that enables us to

[18] The estimates of the structural equation model are direct estimates of effects. In a conventional path model, where all variables are continuous and normally distributed, one can calculate total, direct and indirect effects of each explanatory variable. But ours is not a conventional structural equation model since it includes variables that are categorical, ordered and continuous but non-normally distributed. Thus, decomposition of effects cannot be carried out using standard structural equation conventions.

account for direct and indirect effects as well as for the changing allocation of individuals by outcomes as they progress through the life cycle. This procedure provides us with a solid basis to estimate the magnitude of health selection. By assessing a set of counterfactual scenarios concerning the effects of early health and the pathways by which they may be tied to attainment later in life, we examine the importance of early life health effects for adult health gradients. We use the full set of estimates from the models of the twelve different outcomes reported in Tables 2A and 2B to simulate a population's trajectories through socioeconomic and health status attainment over the life course. Changing selected early life effects to examine their consequences for adult attainment then provides a measure of the differences across counterfactual scenarios conditioned upon the set of social processes that are characterized by the remaining unchanged model estimates.

We use the observed NCDS cohort's distribution of parental characteristics and simulate this population's trajectories through socioeconomic and health status attainment had it been exposed to the regime of transitions and attainment embedded in the structural equations. The birth outcomes of NCDS cohort members reflect predicted values of birth weight given the estimates from our model for low birth weight. Given individuals' resulting allocation to birth weight statuses, they are then subject to the set of estimates from the models for chronic conditions at age 7. We proceed with calculating for each individual a predicted outcome for all outcomes at the different stages of the life course which are reflected in the models. [19] The final product is an average mobility matrix that summarizes the joint distribution of SES and health status by origin and age 46. The matrix contains the conditional probabilities that a cohort member originating (at birth) in a particular social and health status class i will end up in social and health status j at age 46.[20]

The simulations for each MI data set uniquely define $M(k)$, a 12x12 transition matrix associated with MI data set k (k=1,10). To simplify the analysis and presentation of results, we collapse the SES of destination to only three categories ('low' corresponding to SES categories 1 and 2, 'middle' corresponding to categories 3 and 4 and 'high' corresponding to

[19] The simulation incorporates uncertainty in the model estimates by introducing appropriately scaled stochastic errors for each stage of the simulation. Random draws from normal distribution are scaled by the standard deviation of the estimated residual. For ease of interpretation and to minimize extraneous sources of noise in the process, we use estimates from probit functions for binary outcomes and from least squares estimations of education and SES. The resulting set of estimates from these simplifications correspond to the reported estimates in Tables 2A and 2B. To account for the uncertainty in the stochastic terms themselves, mean values are calculated over repeated draws ranging from 100 to 1,000 draws. The optimal number of iterations is determined by the stability of the resulting mean value and varies across the different models.

[20] A unique (and complicating) feature is that the multiple imputation (MI) procedure yields a total of 10 possible transition matrices, not just one. Thus, we proceed to average matrices across these 10 imputed data sets. For each entry in this matrix we produce a point estimate associated with a standard error that is corrected to account for the within- and between-sample variance among the multiple datasets.

categories 5 and 6) and maintain the dichotomous categorization of health status (poor/fair versus all other). Figure 3 illustrates the robustness of both the lagged and contemporaneous SES gradients for health at all three available adult ages (ages 33, 41 and 46) in the observed data to this simplification of the state space to only 6 states. Thus, jointly considering the state of origin(at birth) and of destination (at age 46) leads to a 6x6 matrix. [21] Entry (i,j) of this matrix corresponds to the probability that an individual whose origin is SES-health status i will move to SES-health status j at age 46. These entries summarize the experience that the original cohort would have had from birth to age 46 had their lives been governed by the life course reflected in the set of parameters of the structural equation model. The entries of the matrix correspond to averages over all simulations. To account for the stochastic nature of the MI procedure we average the matrices $M(k)$ and define a super mobility matrix $\sum_k$ itself associated with a matrix of standard errors corresponding to each cell.[22] It is important to note that $M(k)$ is not a steady state mobility matrix as it only reflects the contribution of effects across two generations. Elsewhere we examine the properties of the steady state mobility matrix resulting from simulations of multiple generations (Palloni et al., 2007).[23]

To estimate the size of health selection, we investigate multiple counterfactual scenarios which are distinguished by the relative importance of the paths identified in Tables 2A and 2B which link early health with adult outcomes. We alter selected early life effects and examine the consequences for the simulated class and health distributions. Comparing the differences in the resulting distributions of adult socioeconomic and health status under different scenarios provides a measure of the consequences of such scenarios for adult health gradients. To measure the size of health selection effects for the adult gradient we compare the joint distribution by SES and health status resulting from the model estimates in Tables 2A and 2B with a scenario in which early health effects are set to zero. We also examine the potential magnitude of health selection by comparing the baseline outcomes with a scenario where the estimated effects of early health are amplified by a magnitude of 5. The importance of particular paths which may link early health to adult attainment is evaluated by comparing the baseline outcomes with scenarios where the effects of cognitive skills are set to zero and the effects of class of origin are set to zero.

[21] The simulations are based on models that utilize the full six point scale of socioeconomic status. We only aggregate the original six classes into three in the simulation outcomes.

[22] As above, the standard errors are corrected to account for both within-sample and between-sample variance.

[23] The steady state $M(k)$ provides insights into inter-generational correlates of health, individual attributes and trajectories of attainment and the consequences for population distributions of health and SES. Analysis of the steady state $M(k)$ involves an additional set of assumptions concerning parent matching and differential fertility by SES.

While the adult gradient may be defined with respect to either current or lagged socioeconomic status, the current analysis assigns primacy to the contemporaneous relationship. First, the steeper gradient in the contemporaneous relationship provides a sharper test for our simulation method. Second, using the contemporaneous association also makes our results comparable to conventional research in the area. Third, because the simulation examines counterfactual scenarios relative to a baseline scenario for the relationship between socioeconomic status and health in adulthood it follows that the choosing one or the other should matter little for our inferences. Finally, although focusing on lagged socioeconomic status mitigates endogeneity concerns in the relationship between contemporaneous adult socioeconomic and health status, it is important to note that the evidence for reverse causality may not apply in our case. Reverse causality has proven to be problematic at ages older than the ones we examine here-at a stage in the life cycle where dilution of assets due to illnesses are more likely to occur (Smith, 2004). Also the evidence for reverse causality is somewhat disputable as in some cases it has proven to be quite weak (Adams et al, 2002). Furthermore, the empirical evidence favoring reciprocal links refers to assets and health not to occupation and health, the key classification criteria in the NCDS. In fact, the idea that the occupation of an incumbent at some point in time is the partial result of preceding health status has not received much support even in the British national data (Fox et al., 1982; Fox et al., 1985; Fox and Adelstein, 1978; Blane et al, 1999). Finally, it should be noted that if the contemporaneous correlation we choose to focus on is inflated due to reciprocal causality this will **diminish** the power that health selection effects may have, not exaggerate it.

6 Analysis of results

Table 4 displays a number of statistics pertaining to the simulated mobility matrices and final vectors of population distribution by health status and SES. The statistics included in this table (and in Table 5) are a small subset of those that could be displayed. Since the stories they tell are very similar, we chose statistics that have an intuitive meaning. The rows in the table correspond to alternative scenarios and the columns correspond to statistics. The first column reports the correlation between a continuous measure of predicted SES and indicator of poor health at age 46. The estimate in the top row results when the estimates of low birth weight and chronic conditions at ages 7 and 16 that are reported in Table 2B are operative over the life cycle. The second row reports the correlation in a counterfactual scenario in which the three health effects are not operative. While both estimates are small, there is only a small decline in the correlation between the baseline and the counterfactual.

The second and third columns provide an alternate measure of health selection effects by

emphasizing lifetime mobility in health and SES. The statistic in column 2 corresponds to the conditional probability that an individual born in the highest SES category and healthiest status will end up in the same SES and health status at age 46. In the baseline scenario, this probability is .860 (first row); there is no change when early health effects are not operative and only a modest decrease to .839 when health effects are magnified by a factor of five (third row). In contrast, the probability of remaining in low SES and poor health declines by 42 percent in the absence of health effects and more than doubles when the effects are magnified by 5. This statistic behaves as one should expect: if one eliminates the effects of early health we are de facto eroding one of the sources of downward mobility and, a fortiori, enhancing SES retention at the top of the hierarchy. However, the larger decline in the probability of remaining in the low SES category and poor health is not necessarily suggestive of the importance of early health effects for the adult gradient. An additional set of counterfactual scenarios illustrates how much of this decline may be attributed to dynamics between health and SES during childhood. We undertake two additional sets of counterfactual experiments that examine the interactions of cognitive skills and SES at birth with the effects of early health on the adult gradient. In a scenario where cognitive skills do not exert any effect, the effects of early health on the probability of remaining in the low SES group and poor health are similar to the above result. However, in a scenario where the effects of SES at birth are not operative (rows six and seven), the magnitudes of both probabilities decline by over 50 percent. While there is comparably large percentage decline in these probabilities between these two scenarios, the small magnitude of the probabilities relative to the estimates in rows one and two suggest important interactions between SES at birth and the effects of early health.

The fourth column reports the changes in the SES gradient in health at age 46 that may be attributed to the selection effects of early health. We adopt the estimated coefficient from a logistic model predicting poor health status at age 46 as a function of the predicted continuous measure of SES as a measure of the adult SES gradient in health. When the estimates of low birth weight and chronic conditions at ages 7 and 16 that are reported in Table 2B are operative over the life cycle, this measure of the gradient is equal to -.758. In the counterfactual scenario in which the three health effects are not operative, the estimated logistic coefficient declines by only .004 (-.754). This decline reflects the contribution of the cumulative effects of low birth weight and chronic conditions at ages 7 and 16 over the life cycle on the contemporaneous SES gradient at age 46.

The small magnitude of this change is robust to treating adult SES as an ordered categorical variable. A categorical SES measure was constructed from the predicted continuous measure of adult SES at age 46. There is little change in the estimated logistic coefficients

for indicators of the first five SES categories (the highest SES category is the reference group) between the baseline and counterfactual (zero health effects) scenarios are of similar magnitude. Moreover, the differences in each indicator's coefficient estimates between the baseline and counterfactual scenarios do not show any clear relationship with the level of SES. The small effect of health selection is also robust to the effects of cognitive skills and SES at birth. Comparing rows four and five illustrates a decline in the logistic coefficient of only two percent. While the 8 percent of the adult gradient which may be attributed to health selection in this case is an increase over the prior experiments, it only amounts to a difference of .03 in the odds of reporting poor health.

The main inference from this exercise is straightforward: early health selection effects do not contribute to these measures of the adult SES gradient in health. Even in a scenario where all three health effects are increased by a factor of five, the logistic coefficient increases to only -.780. This increase is equivalent to only 3.5 percent of the gradient when health effects are not operative.

The measures employed in Table 4 may be insufficient to capture the individual mobility in health and SES that may occur conditional upon a given population distribution of health and SES. Figure 4 displays the distribution of adult status by health and the three collapsed categories of SES at age 46. The distributions are shown for the baseline and counterfactual scenarios where health effects are set to zero and increased by a factor of 5. The small shift in this distribution that may be attributed to early health effects illustrates the weak effects of health selection on the overall population distribution of adult attainment. This relative invariance in the distribution also underscores the need for an alternate measure of the gradient that better captures individual mobility in health and SES.

Table 5 reports the probability of poor health conditioned upon current SES. The top panel displays an indicator of the gradient that is intuitively attractive, namely the slope of the conditional probabilities of being in poor health given SES. For the ease of presentation, we present these results for the collapsed set of three SES categories. The first column indicates that the conditional probabilities of bad health are .26, .19 and .14 for the lowest, middle and highest classes, respectively. A linear approximation of the slope of the conditional probabilities is about -.06 or about 12 percent of the maximum slope possible. The counterfactual scenario where health effects are set to zero (column two) has no consequence for this measure of the gradient. Increasing the magnitude of health effects by 5 increases the slope of the gradient to -.075. This represents a 23 percent increase over the gradient slope in the counterfactual scenario of zero health effects but it only implies a .03 additional shift toward the hypothetical maximum gradient in this case (-.5). This is a fairly trivial amount.

Such small health selection effects may be due to measurement difficulties that arise from endogeneity between health status and SES. If endogeneity between health and SES during adulthood becomes increasingly important beginning at age 40, there are many opportunities for bias in estimates of the gradient which may in turn influence the sensitivity of the gradient to our counterfactual experiments. For this reason, the above statistics were defined with lagged gradients adopting SES at age 41 and birth. The results prove robust to a lagged gradient in health defined by SES at age 41. The results with respect to SES at birth are inconclusive due to the weak gradient observed with respect to that period. Figures 5 and 6 illustrate the conditional probabilities for all six SES categories under the baseline and counterfactual scenarios. The small change in the probabilities across all six SES categories is apparent irrespective of the time period of SES that is adopted.

7 Conclusion

Our proposed model of health and socioeconomic attainment over the life course includes explicit and theoretically justified linkages between early traits and adult conditions. The model is a result of joining theoretical insights from economics, epidemiology, educational psychology, demography and social stratification. We utilize a suitable data set and, unlike previous studies, we implement a strategy to mitigate the effects of sample selection through attrition and item non-response by adapting multiple imputation procedures to a longitudinal setting. The data set enables us to observe the uninterrupted experience of an entire cohort for over 40 years. We then use standard procedures to retrieve robust estimates in order to animate a set of social processes that may be evaluated with simulation. Examining four counterfactual scenarios presents the opportunity to measure the size of health selection effects while accounting for the influence of the distribution of individual traits. Analysis of statistics derived from these estimates suggest that though some early traits do have direct effects on adult health and some of them are important antecedents of SES attainment and adult health, the overall relevance of indirect selection pathways as a mechanism accounting for the relation between adult SES and health status is of muted importance.

We caution that these results must be interpreted in the context of the assumptions that are adopted in this exercise. We use our model estimates to predict individuals' low birth weight status in the simulations. Given the relatively small effects of parents' socioeconomic status on low birth weight status that we estimated with the NCDS, there is ample opportunity for the stochastic elements that we build into our simulation to overwhelm these effects. Second, we are unable to account for the effects of unobserved family background characteristics on both child health and adult status attainment. While we might conclude

that such biases would render our model estimates of early child health effects upper bounds of the effects, our estimates of adolescent health effects raise uncertainties about the net direction of bias due to such effects.

However, three further probable sources of bias suggest downward bias in our estimates of health selection which may be large enough to overwhelm any net positive bias that results from the sources discussed above. First, although sample attrition in the NCDS due to child deaths is small (approximately 5 percent of our sample of interest), evidence that these deaths were among the most fragile in our sample suggest downward bias in our estimates of early and adolescent health as well as educational outcomes. Second, our measures of childhood health available in the NCDS are exceedingly coarse. Given more fine-tuned measures, including physical and psychological impairments, it is likely that the relationships could be stronger than those found. This shortcoming is not trivial given the emerging evidence of socio-emotional and mental health effects on schooling attainment (Farkas, 2003; Currie and Stabile, 2006). Third, although adult self-reported health is known to be a good proxy for underlying health conditions, it is far from a flawless measure of health status. In all likelihood we are understating health problems and conditions and thus contracting artificial inequalities that remain concealed and consequently cannot be explained. Note also that in countries such as the UK around 1958, conditions surrounding the lives of the most deprived in the population were in all likelihood vastly superior to the conditions of the most deprived in low income countries, where early malnutrition, stunting, limited growth and development, and insufficiency of micronutrients, are paramount. In these contexts the room for early health status to leave a deep imprint is much larger as is the number of routes through which it may shape adult socioeconomic achievement and adult health inequalities.

References

Adams, P., M.D. Hurd, D. McFadden, A. Merrill, and T. Ribeiro. 2003. "Healthy, Wealthy and Wise? Tests for Direct Causal Paths Between Health and Socioeconomic Status." *Journal of Econometrics.* 112(3-56):57-63.

Almond, D. 2006. "Is the 1918 Influenza Pandemic Over? Long-term Effects of In Utero Influenza Exposure in the Post-1940 U.S. Population" *Journal of Political Economy.*114(4): 672-712.

Almond, D., K. Chay, and D. S. Lee. 2005. "The Costs of Low Birth Weight." *Quarterly Journal of Economics.* 120(3): 1031-1083.

Attanasio, O.P. and C. Emmerson. 2001. "Differential Mortality in the UK." NBER Working Series Paper No. 8241, Cambridge, Massachusetts.

Attanasio, O.P. and H. Hoynes. 2000. "Differential Mortality and Wealth Accumulation." *Journal of Human Resources.* 35(1):1-29.

Banks, M., M. Marmot, Z. Oldfield, and J.P. Smith. 2006. "Disease and Disadvantage in the United States and in England" *Journal of the American Medical Association.* 295:2037-45.

Barker, D. J. P., ed. 1992. *Fetal and Infant Origins of Adult Disease: Papers.* London: British Medical Journal.

Barker, D.J.P. 1995. "Fetal Origins of Coronary Heart Disease." *British Medical Journal.* 311: 171-174.

Behrman J. R., and M. R. Rosenzweig. 2004. "Returns to Birth weight." *Review of Economics and Statistics.* May, 86(2), 586-601.

Black, D., J.N. Morris, C.Smith, P.Townsend, and M.Whitehead. 1988. *Inequalities in Health: The Black Report: The Health Divide.* London: Penguin.

Black, S. E., P. J. Devereux, and K. G. Salvanes. 2005. "From the Cradle to the Labor Market? The Effect of BirthWeight on Adult Outcomes." *Quarterly Journal of Economics.* 122(1):409-439.

Blane, D. 1999. "The Life Course, the Social Gradient, and Health." Pp. 64-80 in *Social Determinants of Health*, eds. M. Marmot and R.G. Wilkinson. New York: Oxford University Press.

Blane, D., G. Davey Smith, and C. Hart. 1999. "Some Social and Physical Correlates of Intergenerational Social Mobility: Evidence From the West of Scotland Collaborative Study." *Sociology.* 33(1):169-83.

Brewer, R. 1986. "A Note on the Changing Status of the Registrar General's Classification of Occupations." *The British Journal of Sociology.* 37(1): 131-140.

Case, A., A. Fertig, C. Paxson. 2005. "The Lasting Impact of Childhood Health and Circumstance." *Journal of Health Economics.* 24(2): 365-389.

Case, A., D. Lubotsky, C. Paxson. 2002. "Socioeconomic Status and Health in Childhood: The Origins of the Gradient." *American Economic Review.* 92 (5): 1308-1334.

Chandola, T., M. Bartley, A. Sacker, C. Jenkinson, and M. Marmot. 2003. "Health Selection in the Whitehall II Study, UK." *Social Science and Med.* 56(10):2059-72.

Cunha, F. and J. J. Heckman. 2008. "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Journal of Human Resources.* 43(4).

Currie, J. 2007. Healthy, Wealthy and Wise: Is There a Causal Relationship between Child Health and Human Capital Development? Unpublished manuscript. Paper presented at the 2007 National Poverty Center Workshop on Early Life Circumstances, Ann Arbor, MI.

Currie, J. and M. Stabile. 2006. "Child Mental Health and Human Capital Accumulation: The Case of ADHD," *Journal of Health Economics.* 25(6):1094-1118.

Drever, F., Whitehead, M. (Eds.) 1997. *Health Inequalities.* London : The Stationery Office.

Duleep, H.O. 1986. "Measuring the Effect of Income on Adult Mortality Using Longitudinal Administrative Record Data." *Journal of Human Resources.* 21(2):238-51.

Duleep, H.O. 1998. "Has the US Mortality Differential by Socioeconomic Status Increased Over Time?" *American Journal of Public Health.* 88(7).

Duncan, G., K.M. ziol-Guest, A. Kalil 2007. Early childhood and adult attainment, behavior and health. Unpublished manuscript. Paper presented at the 2007 National Poverty Center Workshop on Early Life Circumstances, Ann Arbor, MI.

Eikemo, T., C. Bambra, K. Joyce and E. Dahl. 2008. "Welfare state regimes and income-related health inequalities: a comparison of 23 European countries." *The European Journal of Public Health.* 2008 18(6):593-599.

Elo, I.T. and S.H. Preston. 1996. "Educational Differentials in Mortality: United States, 1979-85." *Social Science and Medicine.* 42(1):47-57.

Farkas, G. 2003. "Cognitive Skills and Noncognitive Traits and Behaviors in Stratification Processes." *Annual Review of Sociology.* 29:541-62.

Feldman, J. J., D. M. Makuc, J. C. Kleinman, and J. Cornoni-Huntley. 1989. "National Trends in Educational Differentials in Mortality." *American Journal of Epidemiology.* 129(5):919-33.

Ferri, E., ed. 1993. *Life at 33: The Fifth Follow-Up of the National Child Development Study.* London: National Children's Bureau.

Fitzpatrick, J. and G. Dollamore. 1999. "Examining adult mortality rates using National Statistics Socio-economic Classification." *Health Statistics Quarterly.* 2:33-40.

Fox, A. J., P.O. Goldblatt, and A.M. Adelstein. 1982. "Selection and Mortality Differentials." *Journal of Epidemiology and Community Health.* 36:69-79.

Fox, A. J. and A. M. Adelstein. 1978. "Occupational mortality: work or way of life?" *Journal of Epidemiology and Community Health.* 32: 73-78.

Fox, A. J., P. O. Goldblatt, and D. R. Jones. 1985. "Social Class Mortality Differentials: Artefact, Selection or Life Circumstances?" *Journal of Epidemiology and Community Health.* 39(1):1-8.

Grossman, M. 1972. "On the Concept of Health Capital and the Demand for Health." *Journal of Political Economy.* 80(2):223-55.

Haas, S. 2006. "Health selection and the process of social stratification: the effect of childhood health on socioeconomic attainment." *Journal of Health and Social Behavior.* 47: 339-354

Haas, S. M.M. Glymour and L.F. Berkman. 2007. The Healthy bird gets the worm: childhood health inequality in labor market outcomes over the work career. Unpublished manuscript. Paper presented at the 2007 National Poverty Center Workshop on Early Life Circumstances, Ann Arbor, MI.

Hawkes, D. and I Plewis. 2006. "Modelling non-response in the National Child Development Study," *Journal of the Royal Statistical Society A.* 169(3): 479-491.

Huxley, R., A. Neil and R. Collins. 2002. "Unraveling the fetal origins hypothesis: is there really an inverse association between birthweight and subsequent blood pressure?" *Lancet.* 360: 659-665.

Hummer, R., R. Rogers, and I. Eberstein. 1998. "Sociodemographic Differentials in Adult Mortality: A Review of Analytic Approaches." *Population and Development Review.* 24(3): 553-578.

Illsley, R. 1955. "Social Class Selection and Class Differences in Relation to Stillbirths and Infant Death." *British Medical Journal.* 2:1520-1526.

Illsley, R. 1986. "Occupational Class, Selection and the Production of Inequalities in Health." *The Quarterly Journal of Social Affairs.* 2(2):151-65.

Johnson, R. and R. F. Schoeni 2007. The influence of early-life events on human capital, health status, and labor market outcomes over the life course. Unpublished manuscript. Paper presented at the 2007 National Poverty Center Workshop on Early Life Circumstances, Ann Arbor, MI.

Kitagawa, E. and P.M. Hauser. 1973. *Differential Mortality in the United States.* Cambridge, Massachusetts: Harvard University Press.

Kuh, D. and Y. Ben-Shlomo, Editors. 2004. *A Life Course Approach to Chronic Disease Epidemiology,* 2nd Ed. Oxford: Oxford University Press.

Lauderdale, D. S. 2001. "Education and Survival: Birth Cohort, Period, and Age Effects." *Demography.* 38(4):551-61.

Lynch, S.M. 2003. "Cohort and Life-Course Patterns in the Relationship Between Education and Health: a Hierarchical Approach." *Demography.* 40(2):309-31.

Lynch, S.M. 2006. "Explaining Life Course and Cohort Variation in the Relationship Between Education and Health: The Role of Income." *Journal of Health and Social Behavior.* 47(4): 324-338.

Mackenbach, J.P., A.E. Kunst, A.E.J.M. Cavelaars, F. Groenhof, and J.J. Geurts. 1997. "Socioeconomic inequalities in morbidity and mortality in Western Europe." *Lancet.* 349:1655-9.

J. P. Mackenbach, I. Stirbu, A. R. Roskam, M. M. Schaap, G. Menvielle, M. Leinsalu, A. E. Kunst. 2008. "Socioeconomic Inequalities in Health in 22 European Countries." *New England Journal of Medicine.* 358(23): 2468-2481.

Manor, O., S. Matthews, and C. Power. 2003. "Health Selection: the Role of Inter- and Intra-Generational Mobility on Social Inequalities in Health." *Social Science and Medicine.* 57:2217-27.

Mare, R.D. 1986. "Socioeconomic Careers and Differential Mortality Among Older Men in the United States." Presented at seminar on comparative studies of mortality and morbidity.

Mare, R.D. and A. Palloni. 1988. "Couple Models for Socioeconomic Effects on the Mortality of Older Persons." Center for Demography and Ecology Working Paper 88-07, University of Wisconsin-Madison.

Martorell, R., A. Stein, and D. Schroeder. 2001. "Early Nutrition and Later Adiposity." *Journal of Nutrition.* 131:874S-80S.

Marmot, M., C.D. Ryff, L.L. Bumpass, M. Shipley, and N.F. Marks. 1997. "Social Inequalities in Health: Next Questions and Converging Evidence." *Social Science and Medicine.* 44(6):901-10.

McDonough, P., G. Duncan, D. Williams, and J. House. 1997. "Income Dynamics and Adult Mortality in the United States, 1972-1989." *American Journal of Public Health.* 87(9):1476-83.

Menchik, P.L. 1993. "Economic Status As a Determinant of Mortality Among Black and White Older Men: Does Poverty Kill?" *Population Studies.* 47(3):427-36.

Moore, D.E. and M.D. Hayward. 1990. "Occupational Careers and Mortality of Elderly Men." *Demography.* 27(1):31-53.

Oreopoulos, P., M. Stabile, R. Wald, and L. Roos. 2008. "Short, Medium, and Long-Term Consequences of Poor Infant Health: An Analysis Using Siblings and Twins." *Journal of Human Resources.* 43(1): 88138.

Palloni, A. 2006. "Childhood Health and the Reproduction of Inequalities." *Demography.* 43(4):587-615.

Palloni, A. and D. Ewbank. 2004. "Selection Effects in the Study of Ethnic and Race Differentials in Adult Health and Mortality." *Ethnic Differences in Health in Later Life*, eds. N.A. Anderson, R. Bulatao, and B. Cohen. Washington D.C.: The National Academy Press.

Palloni, A., M. McEniry, R. Wong, and M. Pelez. 2006. "The Tide to Come: Elderly Health in Latin America and the Caribbean." *Journal of Aging and Health.* 18(2): 180-206.

Palloni, A., C. Milesi, R. White, and A. Turner. 2007. "Triggering Inequality and Health Gradients: Do Health Selection Effects Matter?" In J. Surkyn, J. van Bavel, and P. Deboosere, (eds.), *Demographic Challenges for the 21st Century: A state of the Art in Demography.* Brussels: Vrije Universiteit Press.

Palloni, A. and M. Spittel. 2004. "Wealth, Health and Mortality in the US: 1967-1993." Center for Demography and Ecology, University of Wisconsin-Madison.

Pappas, G., S. Queen, W. Hadden, and G. Fisher. 1993. "The Increasing Disparity in Mortality Between Socioeconomic Groups in the United States, 1960 and 1986." *New England Journal of Medicine.* 329(2):103-9.

Plewis, I., Calderwood, L., Hawkes,D. and Nathan,G. 2004. *National Child Development Study and 1970 British Cohort Study Technical Report: Changes in the NCDS and BCS70 Populations and Samples over Time.* London: Institute of Education.

Power, C., C. Hertzman, S. Matthews, and O. Manor. 1997. "Social Differences in Health: Life Cycle Effects Between Ages 23 and 33 in the 1958 British Birth Cohort." *American Journal of Public Health.* 87(9):1499-503.

Power, C., O. Manor, and A. J. Fox. 1991. *Health and Class: the Early Years.* London: Chapman and Hall.

Power, C., O. Manor, A. J. Fox, and K. Fogelman. 1990. "Health in Childhood and Social Inequalities in Health in Young Adults." *Journal of the Royal Statistical Society, Series A..* 153(1):17-28.

Power, C. and S. Matthews. 1997. " Origins of Health Inequalities in a National Population Sample." *Lancet.* 350(9091):1584-9.

Power, C., S. Matthews, and O. Manor. 1998. "Inequalities in Self-Rated Health: Explanations From Different Stages of Life." *Lancet.* 351:1009-14.

Preston, S.H. and I.T. Elo. 1995. "Are Educational Differentials in Adult Mortality Increasing in the United States?" *Journal of Aging and Health.* 7(4):476-96.

Preston, S.H. and P. Taubman. 1994. "Socioeconomic Differences in Adult Mortality and Health Status." Pp. 279-318 in *The Demography of Aging*, eds. L. Martin and S. Preston. Washington, D.C. National Academy Press.

Raghunathan, T. E., J. M. Lepkowski, J. H. van Hoewyk and P. Solenberger. 2001. "A multivariate technique for multiply imputing missing values using a sequence of regression models," *Survey Methodology.* 27, 85-95.

Rogot, E., P. D. Sorlie, N. J. Johnson, and C. Schmitt. 1992. "A Study of 1.3 Million Persons by Demographic, Social and Economic Factors: 1979-1985 Follow-Up." National Institute of Health Publication No. 92-3297.

Rubin, D. B. 1976. "Inference and missing data." *Biometrika.* 63, 581-592.

Rubin, D. B. 1987. *Multiple imputation for nonresponse in surveys.* New York: Wiley.

Smith, J. P. and R. S. Kington. 1997a. "Demographic and Economic Correlates of Health in Old Age." *Demography.* 34(1):159-70.

Smith, J. P. and R. S. Kington. 1997b. "Race, Socioeconomic Status, and Health in Late Life." Pp. 105-62 in *Racial and Ethnic Differences in the Health of Older Americans*, ed. L.G. Martin and B. J. Soldo. Washington DC: National Academy Press.

Smith, J.P. 1999. "Healthy Bodies and Thick Wallets: The Dual Relation Between Health and Economic Status." *Journal of Economic Perspectives.* 13(2):145-66.

Smith, J.P. 2004. Unraveling the SES-Health Connection, *Population and Development Review.* (Supplement) 30:108-132.

Sorlie, P.D., E. Backlund, and J.B. Keller. 1995. "U.S. Mortality by Economic, Demographic, and Social Characteristics: The National Longitudinal Mortality Study." *American Journal of Public Health.* 85(7):949-56.

Sorlie, P., E. Rogot, R. Anderson, N. J. Johnson, and E. Backlund. 1992. "Black-White Mortality Differences by Family Income." *Lancet.* 340:346-50.

Spittel, M.L. 2003. "A Study of Inequalities in Health: The Role of Wealth Differences and Social Context." Thesis, University of Wisconsin-Madison.

Stern, J. 1983. "Social Mobility and Interpretation of Social Class Mortality Differentials." *Journal of Social Policy.* 12(1):27-49.

Tanner, M. and Wong, W. 1987. "The calculation of posterior distributions by Data Augmentation (with discussion)." *Journal of the American Statistical Association.* 82, 528-550.

Todd, P.E. and K. I. Wolpin. 2007. "The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps." *Journal of Human Capital.* 1(1): 91-136.

Vagero, D. and R. Illsley. 1995. "Explaining Health Inequalities: Beyond Black and Barker. A Discussion of Some Issues Emerging in the Decade Following the Black Report." *European Sociological Review.* 11:219-41.

Valkonen, T. 1987. "Social Inequalities in the Face of Death." Plenaires/Congres Europeen De Demographie. Scenances Plenaires. Helsinki, Finland: International Union for the Scientific Study of Population/European Association for Population Studies and Finland Central Statistical Office.

Valkonen, T. 1989. "Adult Mortality and Level of Education: a Comparison of Six Countries." Pp. 142-60 in *Health Inequalities in European Countries*, Fox, John ed. Aldershot, United Kingdom: Gower.

Valkonen, T. 1992. "Social Inequalities in Health in Europe: Evidence and Trends." Paper presented at the Seminario dei Laghi, Baveno, Italy.

Van de Mheen, H., K. Stronks, C.W.N. Looman, and J.P. Mackenbach. 1998a. "Role of Childhood Health in the Explanation of Socioeconomic Inequalities in Early Adult Health." *Journal of Epidemiology and Community Health.* 52:15-19.

Van de Mheen, H., K. Stronks, C.W.N. Looman, and J.P. Mackenbach. 1998b. "Does Childhood Socioeconomic Status Influence Adult Health through Behavioural Factors?" *International Journal of Epidemiology.* 27(3):431-437.

Van de Mheen, H., K. Stronks, J. van den Bos, and J.P. Mackenbach, 1997. "The Contribution of Childhood Environment to the Explanation of Socioeconomic Inequalities in Health in Adult Life: A Retrospective Study." *Social Science and Medicine.* 44(1):13-24.

Vaupel, J., K. Manton, and E. Stallard. 1979. "The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality." *Demography.* 16(3):439-54.

Victora, C.G., L. Adair, C. Fall et al. 2008. "Maternal and child undernutrition: consequences for adult health and human capital," *Lancet.* 371: 340-357.

West, P. 1991. "Rethinking the Health Selection Explanation for Health Inequalities." *Social Science and Medicine.* 32(4):373-84.

White, R. G. 2008. "Sample Selection Correction for Early Health and Later Life Outcomes in England's 1958 Birth Cohort." Working Paper, Center for Demography and Ecology, University of Wisconsin.

Appendix

Figure 1: Health and SES Group Over the Life Course

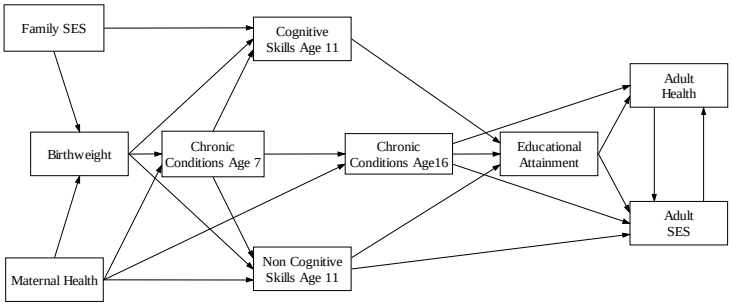


Figure 2: Health Gradient Over the Life Course (Current Socioeconomic Status)

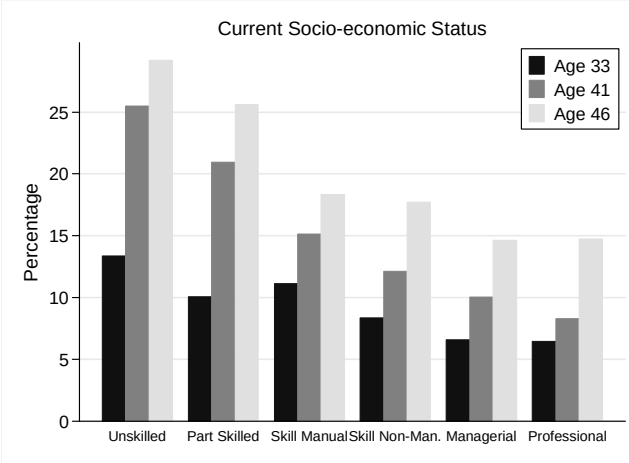


Figure 3: Health Gradient Over the Life Course

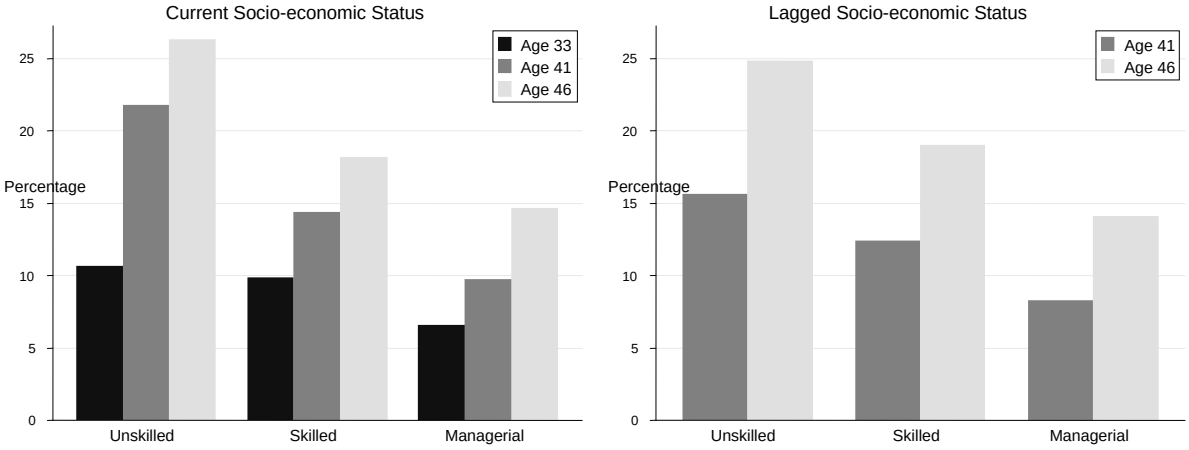


Table 1: Variable Descriptions

Variable	*Description*
SES in Childhood	SES Group of mother's resident husband or partner. Six-categories from 1951 Registrar General's Classification: (1) unskilled manual, (2) semi-skilled manual, (3) skilled manual, (4) skilled non-manual, (5) managerial & technical, (6) professional. If this information was missing at birth because the mother was single or the father had no available information, we use the mother's own SES before pregnancy (about 3% of the cases at birth). For SES measures during childhood, we use the SES of mother's resident husband or partner. If this measure is missing we use the mother's SES.
Obstetrics	(1) Number weeks of gestation. (2) Indicator of twins birth.
Maternal health	(1) Sum of the cohort member mother's retrospective reports of the following conditions: abnormal pregnancy (e.g., preclampsia); and, number of chronic conditions before the child's birth (e.g., diabetes, heart disease, rubella, TB). (2) An indicator of maternal smoking during pregnancy is separately added. (3) A continuous measure of the week number during the pregnancy of first pre-natal medical visit. (4) An indicator of maternal labor force participation during pregnancy.
Early health	Three measures gathered at different points in the cohort member's childhood. (1) Indicator of low birth weight: less than 88 ounces (2,500 grams) at birth. (2) Number of chronic conditions at age 7, reported by a medical practitioner who indicates whether or not the child exhibits each of the following conditions: general motor handicap, disfiguring condition, mental retardation, emotional maladjustment, head and neck abnormality, upper limb abnormality, lower limb abnormality, spine abnormality, respiratory system problem, alimentary system problem, urogenital system problem, heart condition, blood abnormality, skin condition, epilepsy, other central nervous system condition, or diabetes. In the analysis we use a set of indicators of whether the cohort member had 0, 1, 2, or 3 or more chronic conditions. (3) Number of chronic conditions at age 16. In addition to the conditions reported at age 7, a medical practitioner indicated whether the child had any eye, hearing or speech condition.
Cognitive skills	Standardized average of four test scores, measuring four cognitive domains: Verbal, Non-Verbal, Reading Composition and Mathematics; measured when cohort members were 11 years old. Expressed as z-score.
Non-cognitive skills	The natural log of a score of behavioral maladjustment. This score is equivalent to the sum of twelve items representing different aspects of behavioral deviance reported by teachers for cohort members at age 11.
Parent's Education	(1) Mother's age at which schooling was completed. (2) Father's age at which schooling was completed.
Additional Current SES	(1) Family Rents Residence Age 11 (2) Crowding at Home Age 11
Educational attainment	Set of indicators of highest completed education. Indicators represent a five point scale that is derived from the numbers of passed O-level and A-level exams, completed higher education and professional certification. Information collected by schools when cohort members were 20 years old.
Adult SES	Classification of cohort member's own SES , measured at ages 33, 41 and 46. Same 6-category classification as parental SES . When predicting adult SES we treat this variable as an ordered one and use ordered logit models to assess the effects of covariates.
Adult health	Individuals' own self-reported health status at ages 33, 41 and 46. To simplify analyses we grouped the four-point original scale into two categories: poor or fair (=1); and good or excellent (=0). When predicting adult health status we treat this variable as categorical and estimated logistic regressions.

Table 2A. Health and Socio-economic Attainment Over the Life Course: Birth and Adolescence

	LBW	Chronic Conditions Age 7	Cognitive Skills Age 11	LnNon-cog Skills Age 11	Chronic Conditions Age 16	Education
	Logit	NBR	OLS	OLS	NBR	Age 23 OLogit
Low Birth Weight		-0.022	-0.170**	0.052	-0.235	0.065
		(0.156)	(0.075)	(0.088)	(0.319)	(0.173)
Chronic 1 Cond Age 7 [0 Conditions]			0.009	0.014	0.415***	-0.106
			(0.034)	(0.040)	(0.127)	(0.081)
Chronic 2 Cond Age 7			-0.033	0.027	0.357*	-0.073
			(0.055)	(0.065)	(0.202)	(0.130)
Chronic 3+ Cond Age 7			-0.098	0.160*	0.325	-0.176
			(0.081)	(0.095)	(0.297)	(0.196)
Ln Non-Cognitive Skills Age 11						-0.156***
						(0.036)
Cognitive Skills Age 11						0.959***
						(0.046)
Chronic 1 Cond Age 16 [0 Conditions]						-0.006
						(0.118)
Chronic 2+ Cond Age 16						-0.001
						(0.237)
Birth SES Group 1 [Class 6]	-0.241	0.415**	-0.325***	0.130	0.217	-1.478***
	(0.549)	(0.188)	(0.090)	(0.103)	(0.316)	(0.208)
Birth SES Group 2	-0.277	0.271	-0.290***	0.121	0.123	-1.211***
	(0.512)	(0.178)	(0.083)	(0.095)	(0.295)	(0.192)
Birth SES Group 3	-0.053	0.276*	-0.198***	0.020	-0.025	-0.997***
	(0.442)	(0.162)	(0.074)	(0.085)	(0.262)	(0.172)
Birth SES Group 4	0.310	0.064	-0.069	-0.048	0.036	-0.798***
	(0.501)	(0.180)	(0.081)	(0.093)	(0.300)	(0.192)
Birth SES Group 5	-0.410	0.304*	-0.024	-0.059	0.152	-0.516***
	(0.537)	(0.167)	(0.075)	(0.087)	(0.279)	(0.179)
Current SES Group 1 [Class 6][a]		0.003	-0.396***	0.308***	-0.376	-0.594**
		(0.186)	(0.092)	(0.107)	(0.345)	(0.242)
Current SES Group 2		-0.218	-0.345***	0.261***	-0.573**	-0.431**
		(0.153)	(0.074)	(0.086)	(0.276)	(0.198)
Current SES Group 3		-0.056	-0.282***	0.197**	-0.477*	-0.238
		(0.147)	(0.070)	(0.081)	(0.247)	(0.184)
Current SES Group 4		0.083	-0.043	0.085	-0.409	-0.143
		(0.162)	(0.077)	(0.089)	(0.290)	(0.205)
Current SES Group 5		-0.213	-0.102	0.068	-0.671**	-0.144
		(0.152)	(0.069)	(0.080)	(0.262)	(0.185)
Obstetrics	*(included)*					
Maternal Health	*(included)*					
Additional Current SES			*(included)*	*(included)*		
Parent Ed			*(included)*			*(included)*
Observations	4,032	4,032	4,032	4,032	4,032	4,032
R-Squared (Pseudo R-Squared)[b]	0.202		0.132	0.024		0.102
Minimum	0.190		0.125	0.022		0.100
Maximum	0.222		0.141	0.026		0.106
Log Likelihood	-501	-3,741			-1,638	-5,604
Minimum	-476	-3,722			-1,610	-5,571
Maximum	-522	-3,772			-1,667	-5,635

Notes: Estimated models specified in column heading include logit, negative binomial regression (NBR), ordinary least squares (OLS) and ordered logit (Ologit). Standard errors in parentheses. Sample of men from 10 multiply imputed datasets. Reference categories in brackets. Coefficients and standard errors are corrected for within and between sample variances. Model statistics are averaged across datasets. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

[a] Current SES Group for Education equation is class at age 16.

[b] Pseudo R-squared is reported for LBW and Education equations.

Table 2B. Health and Socio-economic Attainment Over the Life Course: Adulthood

	SES Age 33 OLogit	BadHealth Age 33 Logit	SES Age 41 OLogit	BadHealth Age 41 Logit	SES Age 46 OLogit	BadHealth Age 46 Logit
Low Birth Weight	-0.254	0.142	-0.065	0.190	0.004	0.086
	(0.174)	(0.292)	(0.206)	(0.262)	(0.182)	(0.241)
Chronic 1 Cond Age 7 [0 Conditions]	0.045	0.166	-0.239**	-0.092	0.002	0.016
	(0.078)	(0.139)	(0.095)	(0.130)	(0.084)	(0.114)
Chronic 2 Cond Age 7	0.189	0.087	-0.055	-0.124	-0.095	0.105
	(0.124)	(0.224)	(0.153)	(0.209)	(0.135)	(0.178)
Chronic 3+ Cond Age 7	0.017	0.088	-0.176	-0.225	0.375*	-0.276
	(0.179)	(0.319)	(0.216)	(0.312)	(0.198)	(0.288)
Cognitive Skills Age 11	0.264***	-0.127	0.455***	-0.136*	0.275***	0.079
	(0.045)	(0.082)	(0.056)	(0.074)	(0.050)	(0.067)
Ln Non-cognitive skills Age 11	-0.039	0.177***	-0.038	0.092	-0.016	0.091*
	(0.035)	(0.066)	(0.043)	(0.058)	(0.038)	(0.052)
Chronic 1 Cond Age 16 [0 Conditions]	-0.010	-0.015	-0.153	0.076	-0.058	0.089
	(0.113)	(0.207)	(0.138)	(0.183)	(0.121)	(0.161)
Chronic 2+ Cond Age 16	0.182	0.430	0.124	0.119	0.154	0.341
	(0.223)	(0.353)	(0.283)	(0.348)	(0.249)	(0.305)
Education: Less Than 5 O-Levels [None]	0.388***	-0.337*	0.504***	-0.214	0.214*	-0.294*
	(0.105)	(0.177)	(0.128)	(0.165)	(0.114)	(0.153)
Education: 5+ O-Levels Passed	0.157	-0.623***	0.476***	-0.059	0.214*	-0.231
	(0.104)	(0.192)	(0.128)	(0.163)	(0.114)	(0.152)
Education: Any A-levels or Certificate	0.250**	-0.513***	1.124***	-0.336**	0.448***	-0.312**
	(0.095)	(0.165)	(0.119)	(0.153)	(0.106)	(0.140)
Education: University Degree Earned	1.462***	-0.617**	1.988***	-0.472*	1.008***	-0.199
	(0.145)	(0.271)	(0.184)	(0.256)	(0.166)	(0.214)
Birth SES Group 1 [Class 6]	-0.552***	0.296	-0.529**	0.306	-0.641***	0.623**
	(0.190)	(0.389)	(0.231)	(0.317)	(0.209)	(0.296)
Birth SES Group 2	-0.706***	0.506	-0.664***	0.245	-0.447**	0.612**
	(0.175)	(0.363)	(0.214)	(0.299)	(0.194)	(0.279)
Birth SES Group 3	-0.582***	0.350	-0.525***	0.047	-0.481***	0.470*
	(0.155)	(0.338)	(0.189)	(0.274)	(0.172)	(0.255)
Birth SES Group 4	-0.195	0.332	-0.183	-0.394	-0.445**	0.429
	(0.175)	(0.373)	(0.213)	(0.324)	(0.194)	(0.284)
Birth SES Group 5	-0.198	0.171	-0.073	0.117	-0.369*	0.401
	(0.168)	(0.367)	(0.204)	(0.296)	(0.185)	(0.272)
Lag 1 SES Group 1 [Class 6]			-1.209***	-0.031	-6.664***	-0.096
			(0.238)	(0.335)	(0.328)	(0.382)
Lag 1 SES Group 2			-1.405***	0.221	-6.084***	0.504**
			(0.209)	(0.293)	(0.224)	(0.253)
Lag 1 SES Group 3			-1.521***	0.011	-5.261***	0.054
			(0.213)	(0.301)	(0.190)	(0.223)
Lag 1 SES Group 4			-0.609***	0.056	-4.014***	0.416*
			(0.211)	(0.304)	(0.193)	(0.238)
Lag 1 SES Group 5			-0.620***	-0.167	-2.461***	0.067
			(0.217)	(0.322)	(0.164)	(0.206)
Bad Health Age 33			-0.010	1.707***		0.801***
			(0.133)	(0.130)		(0.134)
Bad Health Age 41					-0.240**	1.723***
					(0.102)	(0.112)
Observations	4,032	4,032	4,032	4,032	4,032	4,032
Pseudo R-Squared	0.036	0.028	0.119	0.086	0.252	0.117
Minimum	0.034	0.025	0.109	0.083	0.244	0.115
Maximum	0.037	0.031	0.123	0.091	0.257	0.119
Log Likelihood	-6,374	-1,224	-5,171	-1,428	-4,353	-1,654
Minimum	-6,345	-1,209	-5,098	-1,423	-4,321	-1,650
Maximum	-6,396	-1,243	-5,316	-1,437	-4,398	-1,659

Notes: Standard errors in parentheses. Sample of men from 10 multiply imputed datasets. Reference categories in brackets. Coefficients and standard errors are corrected for within and between sample variances. Model statistics are averaged across datasets. Lag 2 (age 33) is included in models of SES at age 46 and Bad Health at age 46. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3A. Health Age 46. Lagged and Contemporaneous Health and SES.

	SES Covariate Measured at Age 41				*SES Covariate Measured at Age 46*			
	Table 2B	A	B	C	D	E	F	G
SES								
Group 1 [Class 6]	-0.096	0.132	-0.090	0.162	0.442	0.598*	0.442	0.618**
	(0.382)	(0.363)	(0.382)	(0.359)	(0.327)	(0.310)	(0.327)	(0.305)
Group 2	0.504**	0.524**	0.510**	0.523**	0.362	0.433*	0.388	0.474**
	(0.253)	(0.242)	(0.252)	(0.239)	(0.234)	(0.224)	(0.234)	(0.221)
Group 3	0.054	0.079	0.076	0.100	-0.123	-0.073	-0.100	-0.035
	(0.223)	(0.214)	(0.223)	(0.212)	(0.200)	(0.193)	(0.200)	(0.191)
Group 4	0.416*	0.361	0.449*	0.401*	0.068	0.051	0.113	0.116
	(0.238)	(0.229)	(0.238)	(0.226)	(0.225)	(0.217)	(0.224)	(0.214)
Group 5	0.067	0.049	0.084	0.076	-0.084	-0.092	-0.067	-0.070
	(0.206)	(0.198)	(0.206)	(0.196)	(0.185)	(0.179)	(0.184)	(0.177)
Bad Health								
Age 33	0.801***	1.255***			0.803***	1.250***		
	(0.134)	(0.123)			(0.132)	(0.120)		
Age 41	1.723***		1.879***		1.688***		1.845***	
	(0.112)		(0.109)		(0.109)		(0.106)	
Observations	4,032	4,032	4,032	4,032	4,032	4,032	4,032	4,032
Log Likelihood	-1,654	-1,778	-1,673	-1,831	-1,661	-1,782	-1,680	-1,834
Minimum	-1,650	-1,775	-1,669	-1,827	-1,658	-1,779	-1,677	-1,832
Maximum	-1,659	-1,781	-1,678	-1,836	-1,668	-1,786	-1,686	-1,838

Table 3B. Health Age 41. Lagged and Contemporaneous Class

	SES Covariate Measured at Age 33		*SES Covariate Measured at Age 41*	
	Table 2B	A	B	C
Education				
Less Than 5 O-Levels [None]	-0.214	-0.281*	-0.181	-0.255
	(0.165)	(0.160)	(0.164)	(0.158)
5+ O-Levels Passed	-0.059	-0.200	-0.025	-0.170
	(0.163)	(0.157)	(0.161)	(0.155)
Any A-levels or Certificate	-0.336**	-0.435***	-0.274*	-0.377**
	(0.153)	(0.148)	(0.153)	(0.148)
University Degree Earned	-0.472*	-0.601**	-0.445*	-0.565**
	(0.256)	(0.252)	(0.254)	(0.248)
SES				
Group 1 [Class 6]	-0.031	-0.020	0.766**	0.766**
	(0.335)	(0.327)	(0.375)	(0.362)
Group 2	0.221	0.189	0.319	0.329
	(0.293)	(0.288)	(0.289)	(0.281)
Group 3	0.011	0.013	0.154	0.185
	(0.301)	(0.294)	(0.255)	(0.248)
Group 4	0.056	0.025	-0.044	0.043
	(0.304)	(0.297)	(0.281)	(0.273)
Group 5	-0.167	-0.215	0.005	0.046
	(0.322)	(0.316)	(0.242)	(0.237)
Bad Health Age 33	1.707***		1.714***	
	(0.130)		(0.128)	
Observations	4,032	4,032	4,032	4,032
Log Likelihood	-1,428	-1,519	-1,427	-1,518
Minimum	-1,423	-1,512	-1,417	-1,506
Maximum	-1,437	-1,526	-1,435	-1,526

Notes: Standard errors in parentheses. Reference categories in brackets. Sample of men from 10 multiply imputed datasets. All models include the full set of covariates listed in the model for health at age 46 in Table 2B. Coefficients and standard errors are corrected for within and between sample variances. Model statistics are averaged across datasets. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Monte Carlo Simulation Results: SES and Health Mobility Matrix at Age 46

Scenarios	Spearman Mean[a]	Probability Remaining in Good Health High SES	Probability Remaining in Bad Health Low SES	Logistic Regression Coefficient[b]
Effects of cognition and social class of origin unaltered				
Early health effects as estimated	-0.100	0.860	0.158	-0.758
	(.000)	--	--	(.033)
Early health effects set to zero	-0.098	0.860	0.091	-0.754
	(.000)	--	--	(.122)
Early health effects * 5	-0.111	0.839	0.391	-0.780
	(.000)	--	--	(.108)
Effects of cognition set to zero and effects social class of origin unaltered				
Early health effects as estimated	-0.120	0.857	0.175	-1.436
	(.000)	--	--	(.199)
Early health effects set to zero	-0.118	0.863	0.096	-1.413
	(.000)	--	--	(.201)
Effects of cognition unaltered and effects social class of origin set to zero				
Early health effects as estimated	-0.067	0.746	0.062	-1.166
	(.001)	--	--	(.261)
Early health effects set to zero	-0.061	0.754	0.022	-1.063
	(.001)	--	--	(.266)

Notes: [a] p-value in parentheses. [b] Standard error in parentheses. All estimates are mean estimates across 10 multiply imputed datasets. p-values and standard errors corrected for within and between MI dataset variance.

Figure 4: Distribution by SES and Health Status at Age 46

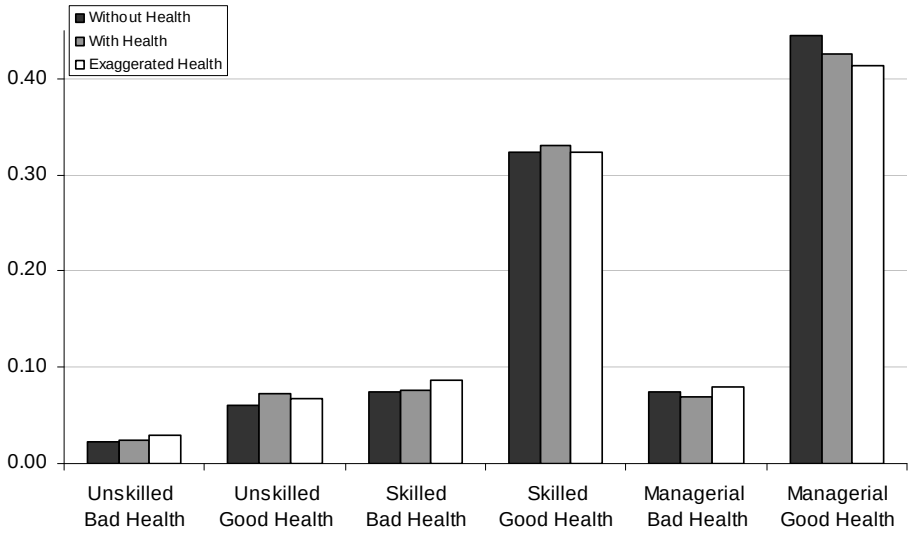


Table 5: Conditional Probabilities of Being in a Good/Bad Health Status Given Social Class

SES and Health Status	Early health effects as estimated		Early health effects set to zero		Early health effects scaled by 5		Cognitive skills effects set to zero		Early health & cognitive skills effects set to zero		SES at birth effects set to zero		Early health & SES at birth effects set to zero	
	Prop.	Ratio	Prop.	Ratio	Prop.	Ratio	Prop.	Ratio	Prop.	Ratio	Prop.	Ratio	Prop.	Ratio
Low SES														
Bad Health	0.259	0.26	0.264	0.26	0.312	0.31	0.289	0.29	0.286	0.29	0.149	0.15	0.15	0.15
Good Health	0.741		0.736		0.688		0.711		0.714		0.851		0.850	
Medium SES														
Bad	0.189	0.19	0.188	0.19	0.211	0.21	0.187	0.19	0.191	0.19	0.114	0.11	0.111	0.11
Good	0.811		0.812		0.789		0.813		0.809		0.886		0.889	
High SES														
Bad	0.140	0.14	0.143	0.14	0.163	0.16	0.136	0.14	0.134	0.13	0.088	0.09	0.088	0.09
Good	0.860		0.857		0.837		0.864		0.866		0.912		0.912	
Slope	-0.060		-0.061		-0.075		-0.077		-0.076		-0.031		-0.031	

Figure 5: Probability of Reporting Fair or Poor Health at Age 46 by SES Age 46

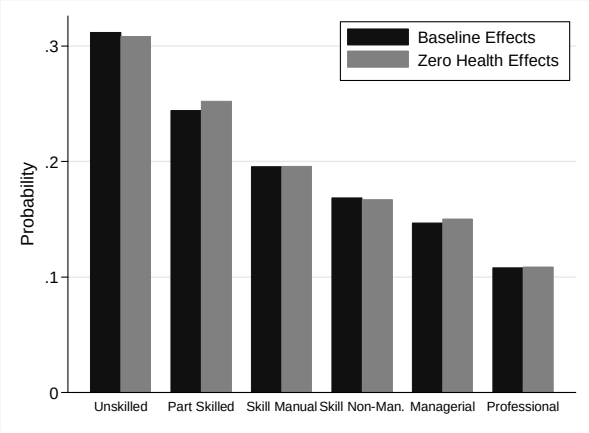
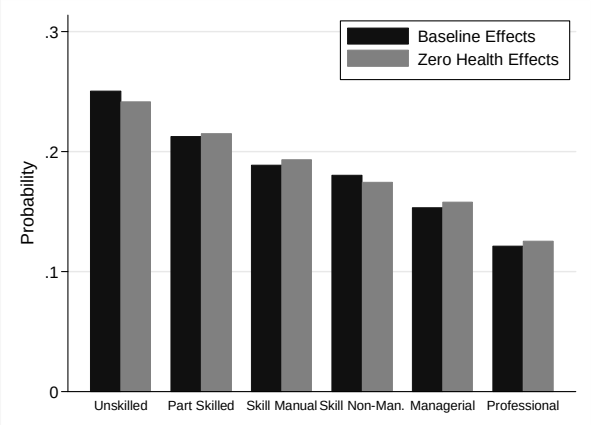
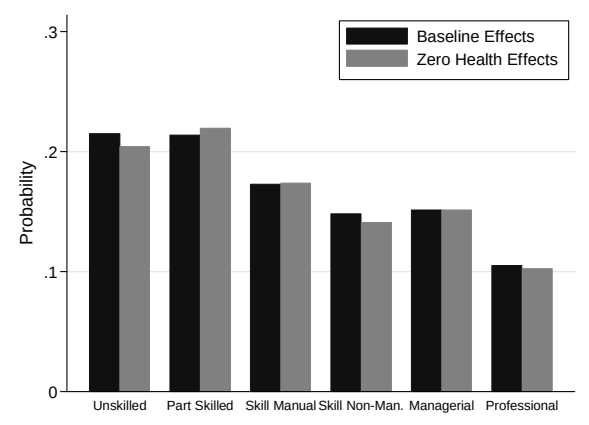


Figure 6: Probability of Reporting Fair or Poor Health at Age 46 with Lags of SES

SES Age 41



SES at Birth



Center for Demography and Ecology
University of Wisconsin
1180 Observatory Drive Rm. 4412
Madison, WI 53706-1393
U.S.A.
608/262-2182
FAX 608/262-8400
comments to: palloni@ssc.wisc.edu
requests to: cdepubs@ssc.wisc.edu