Heterogeneity in the Impact of Type of Schooling on Adult Health and Lifestyle

ANIRBAN BASU[*]
Pharmaceutical Outcomes and Policy Program, Department of Pharmacy, and Departments of Health Services and Economics, University of Washington, Seattle, USA And the National Bureau of Economic Research, Cambridge MA, USA

ANDREW M. JONES
Department of Economics and Related Studies, University of York, Department of Economics, York, UK, and University of Bergen and Centre for Health Economics, Monash University, Victoria, Australia

PEDRO ROSA DIAS
Imperial College Business School, London UK.

ABSTRACT

Using data from a major educational reform in England and Wales, we examine heterogeneity in the long-term impacts of the exposure to different secondary schooling systems, characterized by selective early-tracking system versus non-selective comprehensive schooling, on health outcomes and smoking. We adopt a local instrumental variables approach to estimate person-centered treatment (PeT) effects, thereby recovering the full distribution of individual-level causal effects. We find that the transition from a selective early-tracking system to a non-selective one produced, on a fraction of individuals, significantly increased depression and cigarette smoking. These effects were persistent over time. Cognitive abilities did not moderate the effects, but students with lower non-cognitive skills were most likely to be negatively affected by this exposure. The age-profile of effects indicates that cigarette smoking could be a leading transmission channel of the long-term impact on depression.
Keywords: Educational reform, Health, Lifestyle, Heterogeneity, instrumental variable

The positive health gradient by years of schooling is a particularly robust empirical regularity. Several papers have examined the causal nature of this relationship by exploiting changes in compulsory education laws, such as Lleras-Muney (2005), Kenkel *et al.* (2006), Oreopoulous (2006), Grimard and Parent (2007), Silles (2009), Van Kippersluis *et al.* (2011), Oreopoulous and Salvanes (2011) and Clark and Royer (2013). While the magnitude of the estimated health effects varies considerably among these studies[1], in general, they tend to find positive long-term mean effects of post-compulsory years of education on longevity, health-related behaviors, and outcomes. This body of evidence has inspired some policy interventions targeting young children and adolescents in various parts of the world (e.g., The Quebec Educational Program; Best Start for Kids in King County, Washington).

Nonetheless, as made clear in Clark and Royer (2013), other key aspects of the association between education and health have received much less attention. First, most of the available evidence concerns the effects of length of schooling, rather than the type of that education; this leaves aside all qualitative aspects of an educational system, such as conditions of access, tracking, resources and educational environment. For example, there is a large literature that has looked at the impact of assimilation and also segregation of children from different cultural, socio-economic, and racial backgrounds on their long-term outcomes (Borjas 1985; Fordham and Ogbu 1986; Cutler and Glaeser 1997; Guryan 2004; Card and Rothstein 2007; Echenique et al 2006). More recently, the effects of assimilation of abilities (cognitive and non-cognitive), independent of other factors, have become a focus of attention. Ability tracking within schools, where high

---

[1] These studies use different data, obtain identification from distinct policy reforms and employ different methodologies, hence their results are not directly comparable. Nonetheless, Royer and Clark (2013) find much smaller effects of post-compulsory years of schooling on longevity then Lheras Muney (2005) and Kippersluis (2011). The possible reasons for this, as well as the limitations of the different studies, are discussed in detail in Royer and Clark (2013).

ability students are grouped in separate classes within the same school, have found better results from both high ability and low-ability students. For example, recently Card and Giuliano (2014) found that establishing separate "gifted/high achiever" classrooms for fourth and fifth graders leads to significant achievement gains for non-gifted but high-achiever participants in those classrooms. Beyond ability tracking, mixing of abilities within a school has found mixed results. Kang et al. (2007) exploited a policy experiment in South Korea in 1970 and found that ability-mixing compared to ability-grouping in high school had a positive but statistically insignificant effect on average adulthood earnings. However, this effect was larger for lower ability students and lower, even negative, for higher ability students. Understanding the impact of such mixing of abilities, not only based on cognitive dimensions but also on non-cognitive abilities, on long-term health and lifestyle choices has remained largely unexplored.

In this paper, we try to fill this gap in evidence through three main channels. First, we exploit the implementation of a major education reform in England and Wales that, rather than increasing the school leaving age, exposed pupils of the same age cohort to qualitatively different secondary schooling systems. Geographic variation in the pace that comprehensive schools were introduced in England and Wales means that one group was exposed to a selective early-tracking system in which children took a test to determine whether they were given the opportunity to enroll in an academically intensive secondary school (called 'grammar schools') or else into a less academic secondary school (called 'secondary modern schools). Another group was exposed to a non-selective system, under which testing was abolished and where all pupils, with varying abilities, attended the same type of schools (called comprehensive schools). This allows us to examine the effect of exposure to qualitatively different educational systems, which did vary regarding the curriculum a median student was exposed to and the distribution of

teacher characteristics, and importantly, allowed different levels of assimilation of abilities among the students, on health and lifestyle in adulthood. Non-pecuniary effects of exposure to different types of secondary school systems have been postulated to be sizable in the education sciences and psychology literature (Dupriez et al. 2008 and Perkins and Graham-Bermann 2012) but have received less attention in the economics literature.

Second, we examine the patterns of heterogeneity of effects on health and lifestyle of the exposure to qualitatively different educational systems. Conti et al. (2010, 2011) and Heckman et al. (2014) have used longitudinal data from the UK and the US to characterize heterogeneity health returns to post-compulsory education. Thus, as in Heckman et al. (2014), we adopt an instrumental variables identification strategy and estimate person-centered treatment (PeT) effects (Basu 2014) of selective over non-selective schooling system, thereby recovering the full distribution of the effects and identifying gainers and losers from this policy across the lifecycle.

Third, we examine the roles of cognitive and non-cognitive ability in childhood as possible moderators of the heterogeneity of the estimated effects. Using Dutch data, Bijwaard (2014) has recently shown that cognitive ability can be an important moderator of heterogeneity in the health returns to different lengths of schooling.

We use data from the National Child Development Study (NCDS), which follows a cohort of around 17,000 individuals from their birth in the week of 3rd March 1958. The NCDS cohort-members went through secondary schooling during the 1970s and attended different types of school since their secondary schooling lay within the transition period of the comprehensive education reform that was implemented in

4

England and Wales from the mid-1960s[2]. This reform restructured secondary education by replacing the former system of early tracking of students with a non-selective, comprehensive, educational system. Comprehensive schooling was not implemented simultaneously across the country, and hence only some cohort members attended the selective system of state-funded education, which comprised grammar schools, secondary modern schools and a small number of technical schools. Among members of the NCDS cohort as a whole, 12 percent attended grammar schools at age 16, 25 percent attended secondary moderns, and 57 percent attended comprehensive schools. An additional 6 percent attended private fee-paying schools, independent of the state educational system and reforms[3].

Previous evaluations of the comprehensive schooling reform have focused largely on their direct impact on educational attainment and indirect impact on labour market outcomes (Kerkchoff et al., 1996; Jesson, 2000; Dearden et al., 2002; Galindo-Rueda and Vignoles, 2005, Bonhomme and Sauder, 2011; Burgess et al., 2014). While earlier work that investigated health outcomes used a matching approach and focused on average effects rather than evidence of heterogeneity in responses to the reform (Jones et al., 2011). Here we focus on heterogeneous responses in health-related behaviours and health outcomes.

Our results indicate that there is substantial heterogeneity in the effects of the reform and significant persistence of these effects over time at the individual level. Compared to the selective system of schooling, the newer comprehensive schooling system produced significant negative effects on long-term health and increased smoking among a fraction

---

[2] Comprehensive schooling was introduced in Scotland much earlier than in England and Wales, preventing the use of the Scottish sub-sample as an adequate comparison group for types of school and educational qualifications.

[3] This subsample is thus not used in our analysis.

of individuals. However, the ATE and TT were quantitatively similar and statistically insignificant indicating that cognitive abilities, which were the major drivers for selection into comprehensive schools, did not moderate the effects. Analyzing the PeT effects shows that individuals who had lower non-cognitive skills are most likely to be negatively affected in the long-run from exposure to mixed-ability schools. Our results also show that cigarette smoking could be an important transmission channel of the long-term impact on health outcomes.

1. THE COMPREHENSIVE EDUCATION REFORM

As Kerckhoff *et al.* (1996: 1-2) describe it, following the Second World War, a cycle of educational reforms was initiated in Western Europe. This was characterized by a move away from older selective systems and towards unified, non-selective, systems of comprehensive schooling. From the mid-1960s the comprehensive reorganization took-off in England and Wales, and aimed at promoting equal opportunity in society through an improvement of the quality of the schooling accessible to children from disadvantaged backgrounds. It replaced the selective educational system with a non-selective, comprehensive system of secondary schooling. This policy reform was implemented at different speeds at the local level: some Local Education Authorities (LEAs) implemented it quickly, but others resisted the change. This gradual and uneven transition allowed the two systems to co-exist for a long period.

The NCDS cohort entered state secondary schools at age 11 in 1969 and completed their compulsory schooling at age 16 in 1974; at that point, among those who attended state schools 40 percent of the cohort members still experienced the selective system, while the remaining 60 percent attended comprehensive schools. The NCDS cohort-members

who still experienced the selective system attended either "grammar schools" or "secondary modern" ones. Grammar schools were academically oriented state schools that provided teaching for the entire age range 11-18 and prepared pupils to go on to higher education. Admission into these schools was determined by an exam taken at age 11 (the 'Eleven Plus')[4]. Pupils whose examination score did not permit entry into a grammar school attended either secondary modern schools, which were less academically oriented and covered the ages 11-16 or, in a small minority of cases, vocational schools aimed at providing training and technical apprenticeships[5].

The comprehensive education reform affected schooling in a variety of ways: grammar and secondary modern schools were significantly different regarding their curricula, examinations, and academic environment. They were also very different regarding other qualitative aspects: among the schools attended at age 16 by the NCDS cohort members used in our analysis, 69 percent of grammar schools were single-sex, while only 13 percent of comprehensive schools were single sex. Streaming of classes by academic ability was common in secondary modern schools (42 percent) and comprehensive schools (39 percent) but rare among grammar schools (17 percent). Some comprehensive schools were former secondary moderns (26 percent) or grammar schools (19 percent), with the rest being newly created. Furthermore, the distribution of indicators of quality

---

[4] The earliest comprehensive schools in England and Wales were set up after the Second World War. But the large expansion of comprehensives followed from the policy adopted by the Labour government in 1965. This was implemented through an instruction to local education authorities (LEAs) to plan to convert their schools to comprehensives. In 1970 the new Conservative government, ended the compulsion on local authorities, however, for most the process was well underway. By the mid-1970s most LEAs in England and Wales had converted to the comprehensive system with virtually no secondary modern schools remaining. Many secondary moderns and grammar schools were amalgamated to form comprehensives and a number of new schools were built. As part of the comprehensive reform, the Eleven Plus was abolished as a national examination in England and Wales by the 1976 Education Act. Despite this, the selective system and the existence of grammar schools has persisted in certain areas, such as Kent.

[5] In a few cases, pupils whose CSE grades were sufficient transferred to grammar schools or sixth form colleges to complete their A-levels.

of the school, such as the pupil-teacher ratio and of the average number of expelled students, also differs across types of schools[6].

The issue of the relative advantages of early tracking versus late tracking of students in secondary schools remains a topical education policy issue[7], and the comprehensive reform has received considerable attention in the social sciences (Kerkchoff et al., 1996; Jesson, 2000; Dearden et al., 2002; Galindo-Rueda and Vignoles, 2005, Burgess et al., 2014). Galindo-Rueda and Vignoles (2005) use the NCDS data to investigate the effects of the reform on educational outcomes. Since the type of schooling system attended by each cohort member may be endogenous, they propose an instrumental variables identification strategy, relying on two main instruments: the share of comprehensive schools in a cohort member's LEA in 1974 and the Conservative Party control of the cohort member's LEA (which the authors claim to be negatively correlated with the probability of attending a comprehensive school, but orthogonal to the educational outcomes). However, possible selection into a particular schooling system should be based on the availability of comprehensive schools when the cohort members were 11 years old and ready to enter secondary education, rather than in 1974, when they were 16 old, and this decision had already been made. In addition, as made clear in Pischke and Manning (2006), Conservative areas tended to be significantly richer and better educated than Labour ones, hence the validity of Conservative control of individuals' constituency

---

[6] While the reform affected multiple dimensions of the cohort members schooling experience, evidence from the health literature suggest that the association between many of aspects of the policy (single-sex schools, number of pupils per class, number of pupils expelled from school) and health behaviors and outcomes, is weak both in the short and long run: see for example Jones et al. (2011). Thus, the literature has focused on the effect of the policy through the ability mix.

[7] For example, the OECD evaluation of the results of the Programme for International Student Assessment (PISA) - OECD (2005) - devotes great attention to comparing the relative performance of national education systems with different degrees of selectivity.

as an instrument for the educational system attended is open to question[8] We draw on these concerns as a basis for our identification strategy.

2. EMPIRICAL STRATEGY

As shown by Conti et al. (2010, 2011), individual health returns to schooling are likely to be heterogeneous. In our case, selection into the comprehensive or the selective schooling systems may reflect this type of heterogeneity. Our primary goal is to study such heterogeneity. To do this we apply the local instrumental variables approach of Heckman and Vytlacil (1999) and estimate person-centered treatment effect parameters (Basu 2014) and compare these to conventional IV and OLS estimates. A summary of this method and parameters are given here.

We define a model of potential outcomes and treatment choice using the Neyman-Fisher-Cox-Rubin framework of potential outcomes. Consider two treatment regimes – *comprehensive schools (treated)* state denoted by $D = 1$ and *selective schools (untreated)* denoted by $D = 0$. The corresponding potential individual outcomes in these two states are denoted by Y_1 and Y_0. We assume:

(1) $\qquad Y_1 = \mu_1(X_O, X_U, \vartheta) \text{ and } Y_0 = \mu_0(X_O, X_U, \vartheta),$

where X_0 is a vector of observed random variables, X_U is a vector of unobserved random variables which are also believed to influence treatment selection (they are the unobserved confounders) and ϑ is an unobserved random variable that captures all

---

[8] These concerns have been reinforced in Bonhomme and Sauder (2011). In fact it is difficult to discern a clear pattern in the pace and geographic expansion of the reform, as this was determined simultaneously by a myriad of factors, ranging from existing infrastructure and school size to the evolution of educational theories in the 1960s and 1970s, as well as political affiliation of the respective constituencies.

remaining unobserved random variables. $(X_O, X_U) \perp\!\!\!\perp \vartheta$ and $X_O \perp\!\!\!\perp X_U$ where $\perp\!\!\!\perp$ denotes statistical independence.

Selection into treatment (before the realization of the outcome of interest) is determined by:

(2) $\qquad\qquad\qquad D = 1 \text{ if } \mu_D(X_0, Z) - U_D > 0,$

where Z is a (non-degenerate) vector of observed random variables (instruments) influencing the decision equation but not the potential outcome equations, μ_D is an unknown function of X_0 and Z, and U_D is a random variable that is correlated with X_U and all remaining unobserved random variables influencing choice. By definition, $U_D \perp\!\!\!\perp \vartheta$, which also defines the distinction between X_U and ϑ in (1). Equations (1) and (2) represent the nonparametric models that conform to the Imbens and Angrist's (1994) independence and monotonicity assumptions needed to interpret instrumental variable estimates in a model of heterogeneous returns. As in Heckman and Vytlacil (1999), we can rewrite **(2)** as:

(3) $D = 1 \qquad \text{if } P(X_0 = x_0, Z = z) > V$

where $V = F_{U_D}\left[U_D \mid X_O = x_O, Z = z\right]$, $P(x_0, z) = F_{U_D \mid X_O, Z}\left[\mu_D(X_O, Z)\right]$ and F represents a cumulative distribution function. Therefore, for any arbitrary distribution of U_D conditional on X_O and Z, by definition, $V \sim \text{Unif}[0, 1]$ conditional on X_O and Z. Under regular IV assumptions, Heckman and Vytlacil (1999) show that Marginal Treatment Effects can be identified by:

10

(4) $\dfrac{\partial E_\vartheta(Y \mid X_O = x_0, Z = z)}{\partial p} = E_\vartheta((Y_1 - Y_0) \mid X_O, V = v) = MTE(x_O, v),$

where $Y = DY_1 + (1 - D)Y_0$ is the observed outcomes and $v = P(x_O, z)$.

Basu (2014) extends the LIV methods to identify PeT effects, which, for persons who choose treatment, are defined by:

(5) $E_{X_u \mid X_O, P(Z), D} E_\vartheta(Y_1 - Y_0 \mid x_O, P(z), D = 1) = E(Y_1 - Y_0 \mid x_O, V < P(z))$

$$= P(z)^{-1} \int_0^{P(z)} MTE(x_O, v) \, dv \,.$$

Similarly, the conditional effect for a person who did not choose treatment is obtained by integrating MTEs over values of V greater than p.

Conceptually, a PeT effect is a weighted version of MTEs. For any given individual, the PeT effect identifies the specific margins where that individual may belong given its individual values of X_O, $P(Z)$ and D. It then averages the MTEs over those margins, but not all as in CATE. Therefore, a PeT effect is the *X-Z*-conditional Effect on the Treated for persons undergoing treatment and is the *X-Z*-conditional Effect on the Untreated for persons not undergoing treatment. It is important to note that in the absence of essential heterogeneity, a PeT effect reduces to an individualized CATE where the condition is enforced based on the entire vector of covariate values for an individual. Further technical details can be found in Basu (2014). However, here we provide an explanation of what the PeT effects are and why estimation of the PeT effects are important in this context.

11

2.1 Intuition behind and importance of PeT effects

Suppose that in our example, we presume (only to get the general intuition) comprehensive schooling to be more beneficial, in terms of long-term health, for children with higher test scores in elementary schools and for children who are self-confident. Moreover, choice of treatment is influenced by test scores and the self-confidence of the child. However, our data on children from NCDS do not have a measure of their self-confidence and only contains for each student: test scores, treatment, long-term health outcomes, and a characteristic that all agree is a powerful and valid instrument. Let the instrument be a continuous variable (e.g. distance to the comprehensive school). Greater the distance to the comprehensive less likely will be the enrollment to that school.

In this case, a traditional naïve regression would produce a biased estimate of the average treatment effect (ATE) and also the conditional average treatment effects (CATEs) for children groups with low and high test scores groups due to the endogeneity of the treatment status caused by the missingness of self-confidence levels in the analysis.

A traditional instrumental variable analysis, using a strong and valid IV, will also produce biased estimates for ATE and CATEs due to the presence of *essential heterogeneity*, which suggests that the treatment effects vary over unobserved confounders, which in this case is the self-confidence status.

A local instrumental variable (LIV) approach can be used to overcome these issues when a continuous instrument is available. LIV methods are used to estimate the marginal treatment effects (MTEs) parameters. MTEs are the effects for individuals for whom the

influence of the observed characteristics (tests score and distance to comprehensive school) balance with the influence of the unobserved confounders (self-confidence level) on treatment choice such that they are indifferent to choosing between using the comprehensive and selective schools. To estimate an MTE, LIV methods are used to compare the outcomes of two groups of, say, students with low test scores, where one group is staying at a distance d from the comprehensive school and the other at a distance $d+\square$, $\square$ representing an epsilon (very small) change in distance. Therefore, the two groups are identical in their distribution of self-confidence. Any difference in outcomes between these two groups of students should be attributed to the changes in treatment choices between these two groups, which is only brought about by the difference in distance. However, since the difference in distance is very small, the effect can be attributed to a *margin of students* who were indifferent between two treatment options and were moved to change treatment choice by the small perturbation of the instrumental variable, i.e., distance. For this margin of students, one can *quantify a normalized level of unobserved confounders* as they must balance the observed levels. Similarly, for another dyad of distances, d' and $d'+\square$, one can estimate another MTE, which reflects the causal treatment effect on students at another level of a unobserved confounder. In this way, a full schedule of MTEs can be estimated that vary over the unobserved confounder levels (i.e. self-confidence) given any level of the observed confounder (i.e. test scores). Intuitively, LIV methods estimate these MTEs by estimating the partial derivate of the outcome with respect to the IV-based propensity score (reflecting epsilon changes).

Once MTEs are estimated over the range of observed and unobserved levels, they can then be easily aggregated to form meaningful treatment effect parameters such as the ATE, CATEs, TT and TUT and also study heterogeneity in effects using person-

centered treatment (PeT) effects. The PeT effect for a student in this stylized sample is conditioned not only on that individual's test score but his test score-specific MTEs are also averaged over a distribution of self-confidence statuses that conforms with the individual's observed choice of the schooling system. Thus these are deemed to be personalized effect for this student.

If essential heterogeneity is present, estimation of ATE and CATE would not be possible using traditional IV methods, as explained above. Even in the absence of essential heterogeneity, where one can estimate CATEs for single covariates using traditional IV methods, there is no easy way to develop algorithms that can help understand the combination of covariates that explain treatment effect heterogeneity at the individual level. PeT allows us to develop such algorithm as we demonstrate below. In the context of the policy that we are evaluating, such an algorithm can inform a variety of stakeholders how to group students across types of schooling.

2.2 Identification

Galindo-Rueda and Vignoles (2005) suggest the existence of unobserved individual heterogeneity and selection into types of schooling based on unobserved gains (essential heterogeneity). Hence, traditional instrumental variables approaches would produce biased estimates of mean treatment effect parameters. We follow the LIV approach to estimate PeT effects. Specifically, as a continuous instrument, we use the percentage of comprehensive schools in individual's Local Education Authority in 1969. A similar variable based on 1974 percentages has been used in earlier literature as an instrument to evaluate the reform. However, as explained above, we believe that that 1969 penetration rates of comprehensive schooling provides an even more exogenous variation in schooling system choices since, for this cohort, the transition into either the selective or

14

the comprehensive systems of schooling took place in 1969, when cohort members were 11 years of age, rather than in 1974. Besides directly testing the strength of the instrument, we perform a variety of necessary test on observed risk factors to see whether the instrument can reduce the imbalance in these factors as compared to that observed across school systems. We also adapt the falsification test proposed by Pischke and Manning (2006) to the analysis of health outcomes and estimate the impact of comprehensive schooling on probability of cohort members suffering from at least four childhood morbidities at age 11 (i.e. before exposure to the policy took place) using the 1969 percentage of comprehensive schools in the LEA as an instrument.

2.3 Estimation

To estimate the impact of the comprehensive school system and the individual heterogeneity in these effects, we use local instrumental variable methods (LIV) to compute person-centered treatment (PeT) effects (Heckman and Vytlacil 1999; Basu 2014).

To estimate PeT effects, we follow the following algorithm (Basu, 2014):

1. Run the first stage by regressing the indicator for the comprehensive system (D) against observed factors (X_O) and the instrument (Z) using a probit model and predict the propensity score, $\hat{p}(x_O, z)$, for every individual.

2. Ensure that $\hat{p}(x_O, z)$ has mass at every value (rounded to 0.01) under both comprehensive and selective school systems, dropping any observations that do not.

3. Denote minp $= \min\{\hat{p}(x_O, z)\}$ and maxp $= \max\{\hat{p}(x_O, z)\}$

4. Run the second stage LIV estimand for each binary outcome Y using a probit model with the following specification (determined by a series of goodness of fit tests): $Y = \Phi(\alpha_0 + \alpha_1 \cdot X_O + \alpha_2 \cdot \hat{p} + \alpha_3 \cdot \hat{p} \cdot X_O + \alpha_4 \cdot \hat{p}^2)$, where $\square$ is the cumulative normal distribution function.

5. Numerical integration: For each individual i:

   a. Draw 1000 replications of u~Uniform[minp, maxp]

   b. Compute $d\,\hat{\Phi}(.)/d\hat{p}$ and evaluated it by replacing $\hat{p}(x_O, z)$ with each value of u. So there are 1000 values of $d\,\hat{\Phi}(.)/d\hat{p}$ for each individual i.

   c. Compute D* = $\Phi^{-1}(\hat{p}(x_O, z)) + \Phi^{-1}(1-u)$ also generating 1000 values for each individual i.

   d. Compute the PeT by averaging $d\,\hat{\Phi}(.)/d\hat{p}$ over values of u for which (D* > 0) if D=1, otherwise, by averaging $d\,\hat{\Phi}(.)/d\hat{p}$ over values of u for which (D* ≤ 0) if D=0.

6. Estimated PeT effects provide individualized effects of treatment effects. Mean treatment effect parameters are computed: averaging PeTs over all observation gave ATE, averaging PeTs over over D=1 or D=0 gives the TT and TUT respectively.

We estimate the PeT effects of experiencing comprehensive schooling compared to the selective schooling system on health outcomes (indicators for self-reported excellent health, malaise and long-standing illness) and cigarette smoking (indicator for current smoking) at ages 23, 33 and 42 years separately. We use the estimated PeT effects to study the following:

1. **Differential drop-out**: We examine systematic differences in the distribution of PeT effects at age 23 between those who would continue to participate at age 33

versus those who would drop out before age 33. Similarly, we looked at PeT effects at age 33 based on drop-out behaviors by age 42.

2. **Time dependence of effects**: We explore the dependence of PeT effects on the same outcomes over different ages for the same individuals to study the time dependence or persistence of effects.

4. **Effects on multi-dimensional outcomes**: We compare the dependence amongst the PeT effects of malaise versus long-standing illness for the same individuals to study the multi-dimensional effects of alternative schooling systems. We also repeated similar analysis between the effects on smoking at age 33 years and malaise at age 42 years to establish a potential channel through which alternative schooling systems may affect outcomes.

5. **Heterogeneity**: We studied how the average effects on outcomes among those who did not attend comprehensive schooling systems differed between those who attended secondary modern schools versus grammar schools. This comparison helps to illustrate if exposure to comprehensive schools would have had differential effects on students who went to the different types of schools within the selective system.

6. Finally, we develop a **prediction algorithm** based on individual factors that predict significant PeT effects on malaise and long-standing illness at age 42.

3. DATA

Members of the National Child Development Study (NCDS) cohort are followed from birth, in the week of 3rd March 1958. In the early waves of the Study information was obtained from parents, medical examinations, tests of ability and the child's school; extensive data are thus available on the cohort-members' childhood health, parental background, and educational achievement. The subsequent waves include self-reported

information on social status in adulthood, health-related behavior and a wide range of health outcomes.

Our empirical strategy aims at identifying the impact of the type of secondary schooling experienced by members of the 1958 cohort on adult health and health-related behaviors after controlling for a rich set of pre-treatment factors[9]. These include measures of family socioeconomic status and the local environment during early childhood; measures of childhood health and use of health care and health within the family; and measures of cognitive and non-cognitive skills and social adjustment of the child. Because the comprehensive reform was a secondary education policy, we also condition on characteristics of the individual's primary education, which, in our case, constitute pre-treatment factors. The specific variables used are described below and listed in Table A.1 in Appendix A.

3.1 Pre-treatment factors

Childhood health, genetic endowments, and family background

The NCDS includes extensive information on the childhood characteristics that were shown to be key predictors of health and lifestyle in adulthood in studies such as Case *et al.* (2005) and Currie and Stabile (2004). Following Power and Peckham (1987), we create morbidity indices by aggregating twelve categories of health conditions, which affect the child at ages 7 and 11. We complement this with indicator variables for the occurrence of diabetes, epilepsy, and other chronic conditions among parents and siblings, to account for hereditary conditions in the cohort member's family. Following Carneiro *et al.* (2007), the parental socioeconomic background is proxied by the occupational social class and the years of schooling of cohort members' father. Following Case *et al.* (2005) and

---

[9] As noted by Dearden *et al.* (2002), these reflect early life circumstances, occur prior to secondary schooling, and are not influenced by subsequent educational choices.

Lindeboom *et al.* (2009), we augment this characterisation using data on the incidence of household financial difficulties during the cohort member's childhood and adolescence.

Local area characteristics

Rich information about the cohort members' immediate milieu is available and aggregated at different geographic levels, allowing a detailed demographic and socioeconomic characterization of each individual's local area: electoral constituency level, local education authority level and census enumeration district level (the smallest unit for which census statistics were then available, with an average population of about 500)[10]. Available variables include the percentage of the local population who are unemployed or long-term sick, working women, employed in particular sectors (manufacturing and agricultural), who are in different occupational groups (professional/managerial, other non-manual, skilled manual, semi-skilled, unskilled), owner occupiers, council tenants, non-whites, and immigrants (see Dearden *et al.,* 2002; Galindo-Rueda and Vignoles, 2004; Pischke and Manning, 2006).

Cognitive ability and socio-emotional development

Cognitive skills, socio-emotional development, and personality traits have received growing attention in the literature on human capability formation. In particular, these factors were shown to be key determinants of health and health-related behaviors in recent papers such as Carneiro *et al.*, (2007), Keastner (2009) Conti *et al.* (2010, 2011). The NCDS provides measures of early cognitive ability, and social development collected before respondents began their secondary schooling. Scores of ability tests taken at age 7 and 11 are available on a series of cognitive dimensions: mathematics, reading, copying

---

[10] These small area data are available for the years 1971 and 1981 under a special license, which imposes restrictions on the handling and usage of the data.

designs and general ability. These test scores are highly correlated at the individual level leading to multicollinearity. To avoid this, we follow Galindo-Rueda and Vignoles (2005) and use principal components analysis to construct a single measure of cognitive ability using the first principal component[11].

Serendipitously, the three dimensions of cognitive ability used to construct our index – mathematics, reading and general ability – along with the fact that the index derived from the first principal component gives equal weight to each dimension, mirrors the three elements of the Eleven Plus examination. Cognitive ability score at age 11 can thus be interpreted as a proxy for performance in the Eleven Plus for those who took the examination.

We draw on Carneiro *et al.* (2007) and use the score for the Bristol Social Adjustment Guide (BSAG) as our measure of early social development. This is a measure of problems with social adjustment at age 11: teachers were asked to report whether the child had problems in twelve behavioral domains such as hostility towards children and adults, anxiety, withdrawal, 'writing off' adults, un-forthcomingness, depression, restlessness, acceptance by adults, inconsequential behavior, and other psychological and nervous symptoms (Stott, 1987). One point is attributed to each positive answer; points are then summed to obtain the BSAG social maladjustment score.

3.2 Educational attainment and school characteristics

Detailed information on the cohort-members' educational attainment is available in the NCDS. To exploit this wealth of data, we build a measure of attainment in secondary education with thirteen categories, ordered according to the grades obtained and a

---

[11] At age 11, the first principal component accounts for 85 percent of the joint variation and, strikingly, the weights attached to the three dimensions – 0.583, 0.567 and 0.582 – are virtually identical.

number of passes[12]. In our empirical analysis, we use a simple 0-12 scaling of this variable as a parsimonious measure of educational attainment, although we have also used models with dummy variables for each category to check the robustness of the results.

Schools are characterized using two sets of variables. First, as described in Section 1, the NCDS cohort-members attended different types of schools; this is captured by indicators of the school attended at age 16: secondary modern, grammar, and comprehensive. Second, we use information on characteristics of the school, which have been widely used as standard proxies of the quality of schooling in applied work on education and labor (see, for example, Dearden *et al.,* 2002). These include the teacher/pupil ratio, the ratio of expelled pupils to the total number, and indicators of whether the school was single sex and whether classes were streamed by ability.

Finally, we control for information about the individuals' primary schooling and parental aspirations for their child, which include the number of children in the child's primary school class at age 7 in 1965, whether parents reported that their child was unhappy at school in 1965, and an indicator for whether parents wished their child to continue beyond the minimum school leaving age.

3.3 Outcomes: health and health-related behaviors

We looked three measures of health (depression, long-standing illnesses, and self-assessed health) and one measure of lifestyle behavior (current smoking).

Respondents answer to a series of questions from the Cornell Medical Index Questionnaire, each targeting a particular mental ailment. The number of positive

---

[12] This is variable 'E386' in the NCDS data dictionary; this information was collected in the 1978 Survey of Public Exams, based on a questionnaire sent to the school attended by NCDS respondents at wave 3.

answers given is then used as a malaise/depression score along the lines of Carneiro *et al.* (2007). The malaise score is a measure of psychiatric morbidity (with a special focus on depression), developed at the Institute of Psychiatry from the Cornell Medical Index (Rutter *et al.*, 1970). The NCDS team suggest the use of a severity scale: individuals are considered normal if they score between 0 and 7 points and depressed if they score between 8 and 24 points (Rodgers et al., 1999). In our analysis, we use this scale to define an indicator variable that takes the value one if individuals are depressed (malaise score above 7) and zero otherwise.

Information on the particular medical condition associated with the long-standing illness is available and classified according to the International Classification of Diseases (ICD-10). The most prevalent conditions are illnesses of the musculoskeletal system (25.7 percent), circulatory system (11.8 percent), respiratory system (11.4 percent) and metabolic problems (9.5 percent), of which 70 percent suffer from diabetes. We used an indicator to represent the presence of any long-standing illnesses.

We also look at the probability of individuals reporting excellent health based on self-assessed health on a four-point Likert scale (poor, fair, good, excellent)

The NCDS also documents some of the important health-related behaviors. Classroom peer-effects have been shown to be long-term determinants of these, as shown by Black, Devereux and Salvanes (2008 and 2013). Amongst these health-related behaviors, cigarette smoking has received particular attention. Fletcher (2010 and 2012) provide solid evidence that the adoption of health-related lifestyles such as cigarette smoking and alcohol consumption are partly determined through peer effects during schooling years. Using data from multiple American student cohorts Fletcher (2010) estimates that, on

average, increasing the proportion of classmates who smoke by 10% will increase the likelihood an individual smokes by approximately 3 percentage points, after controlling for the endogeneity of school choice. The long-lasting influence of school-based peer effects on smoking is also documented in European countries. Lundborg (2006) uses rich Swedish data to show that school-class based peer effects in cigarette smoking (and illicit-drug use) are particularly strong. Finally, Etilé and Jones (2011) use French data to show that the effect of schooling reforms can also have long-term effects on cigarette smoking. The NCDS includes self-reported information on the number of cigarettes smoked per day, which we use. These data are only available once respondents are aged 23 and above. Other behavioral data have very high amount of missing data, which precludes us from studying them accurately.

4. RESULTS

4.1 Sample selection, missing values and attrition

Table 1 shows our sample selection and the number of observations lost due to various exclusion criteria. Some NCDS cohort members are excluded from our estimation sample because they attended private schools, which are independent of state education and schooling reforms.

- *Insert Table 1 here -*

More significantly, a large number of observations is lost due to item non-response to parts of the questionnaire that are essential for our analysis or because at some stage they have attrited out of the NCDS. This phenomenon is acknowledged in several papers that feature sample sizes comparable to ours (Dearden *et al.,* 2002; Galindo-Rueda and Vignoles, 2004, 2005; Pischke and Manning, 2006, Lindeboom *et al.*, 2009; Jones *et al.*, 2011). The degree, and repercussions, of non-random attrition has been extensively examined in this earlier literature (for example, Plewlis *et al.*, 2004; Case *et al.*, 2005 and

Lindeboom *et al.*, 2009). Table 2 shows, in the case of our analysis, the comparison between included and excluded observations. The difference in means for most covariates is small and statistically insignificant, with a few notable exceptions. Individuals included in our estimation sample tend to exhibit higher levels of cognitive and non-cognitive ability, to live in slightly more affluent areas and to be less likely to have experienced financial hardship in childhood. We discuss the implications of our results from the included sample in light of these disparities with the excluded sample.

-   *Insert Table 2 here -*

It is also informative to compare the mean differences in covariates between the cohort members in our treatment and control groups; these are reported in Table 3. Average ability levels are higher amongst selective system pupils when compared with their comprehensive counterparts. Also, selective areas tend to be slightly more affluent than comprehensive ones, as identified in Pischke and Manning (2006)[13].

4.2 First stage regression results

Our instrumental variable was strong in predicting selection of comprehensive schooling (F-state = 84, p< 0.0001). Importantly, Table 3 also shows that the differences in covariates are almost eliminated across the median of our instrument;[14] this balancing of observed covariates over the distribution of the instrument is reassuring of its ability to deal with the issues of selection and unobserved heterogeneity.

-   *Insert Table 3 here -*

---

[13] The small number of LEAs where the rate of penetration of comprehensive schools was very close to 100% was systematically different from all the others in terms of observables and was therefore dropped from our analysis.

[14] We also study how these observed characteristics (scaled by their corresponding standard deviation) are balanced across the entire range of the IV and not just above or below the median. The observed levels do not seem to vary more than 0.1SD in most cases (Figure A1 in Appendix A). The largest variation across that entire range of the penetration rate was in mining manufacturing occupation rate, but it was less that 1 SD.

Figure 1 shows the distribution of the values of the propensity score amongst treated and untreated observations. There are only 20 individuals outside the common support, which are dropped before running the second-stage regression.

Also, our IV is not rejected by the falsification test suggested by Pischke and Manning (2006). We estimate the impact of the comprehensive school on the probability of suffering from at least four childhood morbidities at age 11. The estimated ATE, TT, and TUT from this placebo regression approach show null effects: -0.01 (0.08), -0.001 (0.09) and -0.02 (0.07) and they are not statistically significant.

4.3 Second stage regression results: health outcomes

We examine the effect of experiencing the comprehensive schooling system on our outcomes of interest, at three different points of the life-cycle: ages 23, 33 and 42. We also focus on retrieving the PeT effects for each individual in our sample, which are then averaged to calculate the average treatment effect (ATE), the treatment effect on the treated (TT) and the treatment effect on the untreated (TUT). In addition to these overall effects, the PeT effects are used to study the following:

1. **Differential drop-out**: Figure 2 shows the distributions of PeT effects at ages 23 or 33 between those who would continue to participate in the next wave versus those who would drop out before the next wave. We do not observe any systematic differences in these distributions for any of the outcomes.

2. **Time dependence of effects**: Figure 3 illustrates the PeT effects on the probability of depression at age 23 and age 42 for the same individuals. Figure 3 illustrates the

PeT effects on the probability of depression at age 23 and age 42 for the same individuals. Comprehensive schooling, compared to selective schooling, appear to increase the probability of depression at age 42 for 90% of the sample. That is a big jump from 49% at age 23.   We find persistence of the effect of comprehensive schooling over age among individuals who negatively affected beginning at age 23; most of them continue to be negatively affected at age 42. Similar persistence effects were found for other outcomes. However, among individuals who were not affected by comprehensive schooling at age 23, and even may have experienced slight positive effects (i.e., negative PeT effects), most of them would end up experiencing increases in the probability of depression by age 42 attributable to comprehensive schooling (North-west quadrant in Figure 3). It is plausible that realization of ability gap, which may have been masked under the comprehensive schooling system, have caught up with these individuals at a later age. However, this is only conjecture as we do not have sufficient data to pinpoint the channels through which such effects are working.

- *Insert Figure 3 here -*

4. **Effects on multi-dimensional outcomes**: Table 4 shows the effect of comprehensive schooling on the probability of cohort members suffering from depression, having a long-standing illness and self-assessed excellent health and smoking. In general, we do not find significant evidence of essential heterogeneity for any of the outcomes. The IV-OLS average effects are all similar to the ATEs estimated via LIV methods.   Over the ages, the effect of comprehensive schools seems to have increased the probabilities of depression and long-standing illnesses and decreased the probability of excellent self-assessed health, but none of these effects reach significance at the 5% level.   The absence of essential heterogeneity does not signify the absence of heterogeneity in effects since the observed risk

factors may still moderate these effects. Indeed, we see that the point estimates of the unconditional TT and TUT effects are different from each other, especially for the probabilities of depression and long-standing illnesses.

- Insert Table 4 here -

5. **Heterogeneity**: Figure 4 shows the PeT effects of comprehensive schooling on the probability of depression among untreated individuals, i.e. those who experienced the selective system of education. Interestingly, the mental health losses associated with comprehensive schooling would be similar amongst cohort-members who attended secondary modern schools (the less academically intensive schools) or the grammar schools, highlighting that cognitive abilities, which distinguish these two groups of students, may not be a driving factor for these effects.

- Insert Figure 4 here -

However, if we perform univariate variance decomposition of the PeT effects, which estimates how much of the overall variation in the PeT effects on the probability of depression at age 42 by each of the observed baseline factors, we find six univariate factors that each explain at least 5% of the variance in the PET effects (Table 5). Both cognitive and non-cognitive scores in childhood are identified amongst these factors, indicating univariate CATEs, as typical in most subgroup analysis, would have identified cognitive scores as a factor explaining heterogeneity in the effects of comprehensive schooling over selective schooling. That would be at odds with our previous evidence and shows the value of developing a prediction algorithm using the PeT effects.

Insert Table 5 here -

6. **Prediction algorithm.** Since it is important to identify the gainers and losers from comprehensive schooling over selective schooling. we use a stepwise backward logistic regression approach on the PeT effects, with very conservative inclusion criteria of p-value < 0.0001, to jointly identify the baseline factors that are most influential in predicting the individuals for whom comprehensive school would have significantly increased (at p <=0.10) in probability of depression by age 42. The results, presented in Table 6, show the key factors that identify these individuals do not include cognitive ability, in line with what we found in Figure 4. Nonetheless, children with one point increase in their social adjustment scores (which increase in non-cognitive difficulties) are at a 10% greater risk of a significant increase in the probability of depression at age 42.

- Insert Table 6 here -

4.4 Transmission channels: quality of schooling, cigarette smoking and health

To explore whether the effect of comprehensive schooling on long-term health outcomes such as depression and long-standing illnesses could have been mediated through its influence on lifestyle, we explore the association of PeT effects on smoking behavior at age 23 with the PeT effects on depression and long-standing illness at age 42 for the same individuals. Figure 5 shows that there appear to be strong positive correlations in both cases. The positive PeT effects of comprehensive schooling on smoking track positive effects on probabilities of suffering from depression (rho = 0.47; 95% CI: 0.13, 0.81) and long-standing illness (rho = 0.27; 95%CI: -0.10, 0.64) at age 42. In contrast, we find that the correlation between the PeT effects for smoking and long-standing illness or smoking and depression are close to zero at age 23 (rho = 0.06 and 0.09 respectively). This contrast in correlations corroborates the role of cigarette smoking

as a transmission mechanism of the impact of quality of schooling on health outcomes, especially depression.

- *Figure 5 here* –

5. DISCUSSION

We studied the effect of exposure to selective versus non-selective schooling system among secondary school students by exploring the implementation of the comprehensive education reform in England and Wales which, rather than increasing the school leaving age, exposed pupils of the same age cohort to qualitatively different secondary schooling systems. We estimated the effects on long-term health and lifestyle outcomes within a causal framework and explored heterogeneity in these effects.

Similar to the story that Heckman et al. (2014) find regarding the effects of length of schooling health outcomes, especially mental health, our analysis shows that the 'devil' is in the distribution (Conway and Deb 2005). In our work, exposure to comprehensive schooling system, compared to a selective schooling system in adolescence, exposes students to a different set of curriculum and distribution of teacher quality and also a mix of peers with a much broader range of cognitive and non-cognitive skills. Our results indicate that the average impacts of such exposure led to an increased smoking prevalence, that peaks at age 33, and increase probability of depression and long-standing illness at ages 33 and 42. However, the average effects, seemingly large, are not statistically significant. In fact, we do not see differences in the average treatment effect and the effect of the treated, where selection into comprehensive schools was driven primarily cognitive ability levels. However, we do find considerable variation in

individual treatments effects (in line with Heckman et al. 2014), and the magnitude of these effects are persistent within individuals as they age. We find that if all children in the NCDS cohort were sent to comprehensive schools, about 50% of them would have significant ($p < 0.10$) increases in the probabilities of long-standing illness or depression at some point in their adulthood (up to age 42). Those who would be significantly worse off on depression score at age 42 due to comprehensive schools are more likely to be men and those who had lower non-cognitive abilities at age 11. These results align well the recent literature on mixing abilities. Card and Giuliano (2014) show that students with lower non-cognitive skills (Plan B students) would not benefit much from exposure to high ability kids. In fact, we show that they may get hurt in the long run. Similar results were also found by Kang et al. (2007) study, which exploits a policy experiment in South Korea[15].

One limitation of our analysis is that these results are contingent upon analyzing one cohort. However, this one cohort, which is the population of people born in one specific week in the entire UK, is the only of the large British cohorts whose schooling years are in the transition period of the comprehensive education reform and lines up well with their entry age to secondary schools and the timing of the reform. Thus, it enables us to study these effect right from entry, which is important because the UK government has recently announced plans for re-introducing a system of selective education based in the selective system abolished by the reform and highlights that the debate on the relative

---

[15] Our results also align well with the earlier education literature on the comprehensive education reform. Galindo-Rueda and Vignoles (2005) evaluate this reform in terms of educational attainment at age 16. Their estimates indicate that pupils in the top 20% of the cognitive ability distribution do significantly worse under the non-selective (mixed ability) system, relative to the selective one. Also low-ability students did not improve their scores in the mixed ability system, leading the authors to conclude that they did not benefit from having more able peers.

merits of comprehensive versus selective schooling is still a topical issue[16]. Our paper

provides evidence that is relevant to this debate[17].

Our results can also provide useful evidence to the ongoing debates, in the US and other

countries, on the usefulness of grouping students around abilities and targeting of the

curriculum. In fact, much of these debates have centered around grouping based on

cognitive abilities (e.g. designated gifted schools in the US). More discussions and

research are needed to precisely delineate and fully understand the effects of grouping

students based on non-cognitive abilities.

---

[16] The UK Government plans for the re-introduction of early tracking selective schooling includes bringing back the grammar school that were removed by the comprehensive schooling reform. This issue has dominated the front pages in the UK press as seen, for example, here:
https://www.theguardian.com/education/grammarschools

[17] As mention above, a second limitation of this analysis (which is common to the entire literature on this reform) is its inability to fully disentangle all the channels through which it affected the behavior and outcomes of the NCDS cohort members. This would require sources of independent in its multiple dimensions, which is not available.

REFERENCES

Basu A. 2014. "Person-Centered Treatment (PeT) effects using instrumental variables: An application to evaluating prostate cancer treatments." *Journal of Applied Econometrics* 29:671-691.

Bijwaard, G., Kippersluis, H. and Veenman, J. 2013. "Education and health: the role of cognitive ability". *IZA DP7648*

Bonhomme, S. and Sauders, U. 2011. "Recovering distributions in difference-in-differences models: a comparison of selective and comprehensive schooling." *Review of Economics and Statistics*, 93(2): 479 – 494.

Borjas, G. (1985) Assimilation, Changes in Cohort Quality and the Earnings of Immigrants. *Journal of Labor Economics*, 3(4), 463–489.

Burgess, S. Dickson, M. and Macmillan, L. 2014. "Selective schooling systems increase inequality". Institute of Education working paper 14-09.

Card D, Giuliano L. 2014. Does gifted education work? For which students? NBER Working Paper # 20453. http://www.nber.org/papers/w20453.

Card A, Rothstein J. (2007) Racial segregation and black-white test score gap. Journal of Public Economics XCI: 2158-2184.

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

Carneiro, P. Crawford, C. and Goodman, A. 2007. "The impact of cognitive and non-cognitive skills on later outcomes." CEE Discussion Papers.

Clark, D. and Royer, H. 2013. "The Effect of Education on Adult Mortality and Health: Evidence from Britain." *American Economic Review*, 103(6): 2087-2120.

Conti, G., Heckman, J. and Urzúa, S. 2010. "The Education-Health Gradient" *American Economic Review* 100(2), pp. 234-38.

Conti, G., Heckman, J. and Urzúa, S. 2011. "Early endowments, education and health". Human Capital and Economic Opportunity Working Group Working paper 2011 011 – University of Chicago.

Conway, K.S. and Deb, P. 2005. Is prenatal care really ineffective? Or, is the 'devil in the distribution? *Journal of Health Economics* 24(3): 489 – 513.

Cunha, F. and Heckman, J. 2007. "The Technology of Skill Formation." *American Economic Review* 97(2), pp. 31-47.

Cunha, F. and Heckman, J. 2008. "Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation." *Journal of Human Resources*

Cunha, F., J. Heckman, J. and Schennach, S. 2010 "Estimating the Technology of Cognitive and Noncognitive Skill Formation", *Econometrica* 78(3), p.p. 883-931.

Currie, J. and Moretti, E. 2003. "Mother's education and the intergenerational transmission of human capital: evidence from college openings." *Quarterly Journal of Economics*, 118:1495–1532.

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

Cutler D, Glaeser E. (1997) Are ghettos good or bad? Quarterly Journal of Economics CXII: 827-872.

Department of Education and Science 1971. "Statistics of Education – 1969", Her Majesty's Stationary Office, London.

Dearden, L. Ferri, J. And Meghir, C. 2002. "The effect of school quality on educational attainment and wages." *Review of Economics and Statistics.* 84:1-20.

Dupriez V, Dumay X, Vause A. How do school systems manage pupil's heterogeneity? *Comparative Education Review* 2008; 52(2): 245-273.

Echenique F, Fryer R, Kaufman A. (2006). Is school segregation good or bad? American Economic Review XCVI: 265-269.

Etilé, F. and Jones, A. M. 2011. Schooling and smoking among the baby boomers – An evaluation of the impact of educational expansion in France. *Journal of Health Economics* 30:4811-831.

Fletcher, J. 2010 Social Interactions and Smoking Decisions: Evidence Using Multiple Cohorts, Instrumental Variables, and School Fixed Effects. *Health Economics* 19(4) 466-484

Fletcher, J. 2012.Peer influences on adolescent alcohol consumption: evidence using an instrumental variables/fixed effect approach. *Journal of Population Economics*, 25(4) 1265-1286

Fordham, S. and Ogbu, J.U. (1986) Black Students' School Success: Coping with the Burden of 'Acting' White. *The Urban Review*, 18(3), 176–206.

Galindo-Rueda, F. and Vignoles, A. 2004. "The heterogeneous effect of selection in secondary schools: understanding the changing role of ability", IZA Discussion Paper No. 1245.

Galindo-Rueda, F. and Vignoles, A. 2005. "The declining relative importance of ability in predicting educational attainment." *Journal of Human Resources,* 40: 335-353

Grimard, F. and Parent, D. 2007. "Education and smoking: Were Vietnam War draft avoiders also more likely to avoid smoking?" *Journal of Health Economics*, 26: 896-926.

Guryan J. (2004) Desegregation and black drop-out rates. American Economic Review XCIV: 919-944.

Heckman, J.J., Humphries, J.E., Veramendi, G. and Urzúa, S. 2014. "Education, health and wages". NBER working paper 19971.

Heckman, J.J., Urzúa, S. and Vytlacil, E. 2006 "Understanding instrumental variables in models with essential heterogeneity" *Review of Economics and Statistics*, 88(3): 389-432.

Heckman, J.J., and Vytlacil E.J. 1999. Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the National Academy of Sciences*, 96(8): 4730-34.

Imbens G. and Angrist J. 1994. Identification and estimation of local average treatment effects. *Econometrica*, 62(2): 467-475.

Jesson, D. 2000. "The comparative evaluation of GCSE value-added performance by type of school and LEA." University of York discussion paper No. 2000/52.

Johnson, R. 2010 "The Health Returns to Education Policies: From Preschool to High School & Beyond" *American Economic Review*, 188-94.

Jones A.M., Rice, N. and Rosa Dias, P. 2011. "Long-term effects of cognitive skills, social adjustment and schooling on health and lifestyle: evidence from a reform of selective schooling". *Journal of Human Capital* 5(3) 342-76.

Lundborg, P. 2006. Having the wrong friends? Peer effects in adolescent substance use. *Journal of Health Economics* 2006 Mar;25(2):214-33.

Kang, C., Park, C. and Lee M-J. 2007. Effects of ability mixing in high school on adulthood earnings: quasi-experimental evidence from South Korea. Journal of Population Economics 20: 269-297.

Keastner, R. 2009. "Adolescent cognitive and non-cognitive correlates of adult health." Technical report, National Bureau of Economic Research.

Kenkel, D., Lillard, D. and Mathios, A. 2006. "The roles of high school completion and GED receipt in smoking and obesity", *Journal of Labor Economics*, 24: 635-660.

Kerkchoff, A.C. Fogelman, K., Crook, D. and Reeder, D. 1996. "*Going comprehensive in England and Wales: A Study of Uneven Change.*" London: The Woburn Press.

Lindeboom, M., Llena-Nozal, A. and Van der Klaauw, B. 2009. "Parental education and child health: evidence from a schooling reform." *Journal of Health Economics*, vol. 28, pp. 109-131.

Lleras-Muney, A. 2005. "The relationship between education and adult mortality in the United States." *Review of Economic Studies*, vol. 72, pp. 189-221.

Oreopoulos, P. 2006. "Estimating average and local average treatment effects of education when compulsory schooling laws really matter." *American Economic Review*, vol. 96, pp. 152-175.

Oreopoulous, P. and Salvanes, K. 2011."Priceless: The Nonpecuniary Benefits of Schooling" *Journal of Economic Perspectives* 25(1): 159-84

Perkins S, Graham-Bermann S. Violence exposure and the development of school-related functioning: mental health, neurocognition and learning. *Aggressive and Violent Behavior* 2012; 17(1): 89-98.

Pischke, S. and Manning, A. 2006. "Comprehensive versus selective schooling in England and Wales: what do we know?" IZA DP No. 2072.

Plewis, I. and Calderwood, L. et al. 2004. "Changes in the NCDS and BCS70 populations and samples over time". London, Centre for Longitudinal Studies.

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

Rodgers, B., Pickles, A., Power, C., Collishaw, S. and Maughan, B. 1999 "Validity of the Malaise Inventory in general population samples." *Social Psychiatry and Psychiatric Epidemiology*, vol. 34, pp.333-341.

Rutter, M., Tizard, J. and Whitemore K. 1970. "Education, Health and Behaviour", London.

Silles, M. 2009. "The causal effect of education on health: evidence from the United Kingdom", *Economics of Education Review*, vol. 28, pp. 122-128.

Stott D. H. 1987. "The social adjustment of children: Manual to the Bristol Social Adjustment Guides." London: Hodder and Stoughton

Van Kippersluis, H., O. O'Donnell, and E. van Doorslaer (2011). "Long Run Returns to Education: Does Schooling Lead to an Extended Old Age?" Journal of Human Resources 46(4): 695-721.

Figure1: *Distribution of the propensity score P(X, Z): observations within and off common support*