

**“THE EDUCATION-HEALTH GRADIENT”**

**DATA APPENDIX**

Gabriella Conti  
University of Chicago

James J. Heckman  
University of Chicago  
and University College Dublin

Sergio Urzúa  
Northwestern University  
and IZA

January 9, 2010

# 1 The British Cohort Study (BCS70)

We use data from the British Cohort Study (BCS70), a survey of all babies born (alive or dead) after the 24th week of gestation from 00.01 hours on Sunday, 5th April to 24.00 hours on Saturday, 11 April, 1970 in England, Scotland, Wales and Northern Ireland.<sup>1</sup> There have been six follow-ups so far to trace all members of the birth cohort: in 1975, 1980, 1986, 1996, 2000, and 2004. We draw information from the birth survey, the second sweep (age 10) and the fifth sweep (age 30). After removing children born with congenital abnormalities and non-whites (or those with missing information on ethnicity), and deleting responses with missing information on the covariates, we are left with a sample of 3,777 men and 3,620 women.

## 1.1 Schooling and Post-Schooling Outcomes

The outcomes considered in our model are:

- **Schooling.** Our schooling measure is a dummy variable indicating whether or not the individual stayed on in school after reaching the minimum school-leaving age. For the individuals in our data, the minimum school-leaving age was 16 years.
- **Labor Market Outcomes.** We analyze two labor market outcomes: (log) hourly wages and full-time employment status. Both are measured at age 30.
- **Healthy Behaviors.** We analyze three healthy behaviors, all measured at age 30.
  - ever used cannabis: this variable takes the value 1 if the individual reports having used cannabis by age 30;
  - daily smoking: this variable takes the value 1 if the individual smokes cigarettes every day;
  - regular exercise: this variable takes the value 1 if the individual does any regular exercise.
- **Health.** We include three variables characterizing individual’s health status by age 30.
  - self-reported poor health: this variable takes the value 1 if the individual reports his/her health to be generally “fair” or “poor”;

---

<sup>1</sup>The original name of the data was the British Births Survey (BBS), sponsored by the National Birthday Trust Fund in association with the Royal College of Obstetricians and Gynecologists.

- obesity: this variable is constructed in the standard way as having a BMI>25 (for females) or a BMI>30 (for males), where the BMI is weight in kilograms divided by height in meters squared.<sup>2</sup>
- depression: this variable takes the value 1 if the individual is categorized as depressed; it is measured using the the Malaise Inventory (Rutter et al., 1970), which includes 24 ‘yes-no’ items which cover emotional disturbances and associated physical symptoms.

## 1.2 Measurement System

As indicators of cognitive ability we use the following seven test scores administered to the children at age ten:

- the Picture Language Comprehension Test: this is a new test specifically developed for the BCS70 on the basis of the American Peabody Picture Vocabulary Test and the English Picture Vocabulary Test; it covers vocabulary, sequence and sentence comprehension.
- the Friendly Math Test: this is a new test specifically designed for the BCS70; it covers arithmetic, fractions, algebra, geometry and statistics.
- the Shortened Edinburgh Reading Test: this is a shortened version of the Edinburgh Reading Test, which is a test of word recognition particularly designed to capture poor readers; it covers vocabulary, syntax, sequencing, comprehension, and retention.
- the four British Ability Scales: they measure a construct similar to IQ, and include two verbal scales (Word Definition and Word Similarities) and two non-verbal scales (Recall Digits and Matrices).

We performed a preliminary factor analysis of these measurements. Both Velicer (1976) minimum average partial correlation criterion and Kaiser (1960) eigenvalue rule suggested to retain one component.

As measurements of noncognitive ability we use six scales collected at age ten:

---

<sup>2</sup>We use a different threshold for males and females as the difference between high- and low-educated females is barely statistically significant if we use as threshold BMI>30.

- the locus of control (caraloc) scale: this is administered to the child and includes sixteen items which measure whether an individual’s locus of control is external or internal;
- a perseverance scale: this is administered to the teacher, who answers to the question “How much perseverance does the child show in face of difficult tasks?” on a scale from 1 to 47;
- a cooperativeness scale: this is administered to the teacher, who makes an estimate of how cooperative is the child with his peers, on a scale from 1 to 47;
- a completeness scale: this is administered to the teacher, who assesses “The child completes tasks which are started”, on a scale from 1 to 47;
- an attentiveness scale: this is administered to the teacher, who assesses “Child pays attention to what is being explained in class”, on a scale from 1 to 47;
- a persistence scale: this is administered to the teacher, who assesses “Child shows perseverance, persists with difficult or routine work”, on a scale from 1 to 47.

We performed a preliminary factor analysis of these measurements. Both [Velicer \(1976\)](#) minimum average partial correlation criterion and [Kaiser \(1960\)](#) eigenvalue rule suggested to retain one component. The Cronbach’s alpha reliability coefficient for this scale is 0.81.

Finally, we use the following measures for the health endowment, recorded at age ten:

- height (recorded during the medical visit);
- head circumference (recorded during the medical visit);
- father’s height;
- mother’s height.

We performed a preliminary factor analysis of these measurements. Both [Velicer \(1976\)](#) minimum average partial correlation criterion and [Kaiser \(1960\)](#) eigenvalue rule suggested to retain one component.

### 1.3 Observed Characteristics

We include the following set of covariates in both the measurement system and in the outcome equations:

- mother’s age at birth;
- a dichotomous variable for mother’s education at birth (whether or not the mother continued education beyond the minimum school-leaving age);
- a dichotomous variable for father belonging to high social class at birth (Social Class I, II or III Non Manual);<sup>3</sup>
- a categorical variable for total gross family income at age 10;<sup>4</sup>
- a dichotomous variable for whether the child lived with both parents since birth until age 10;
- parity;
- the number of children in the family at age 10.

We also include child’s weight in the measurement equation for child’s height and head circumference, and mother(father)’s weight in the measurement equations for maternal(paternal) height. The schooling choice model also includes as covariate the gender-specific seasonally-adjusted rate of unemployment-related benefit claims (the claimant count) as observed in January 1986.

## References

- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement* 20(1), 141–151.
- Rutter, M., J. Tizard, and K. Whitmore (1970). *Education, health and behaviour*. Longmans London.

---

<sup>3</sup>The BCS70 uses the Registrar General’s classification for measuring social class (SC). Social class I includes professionals, such as lawyers, architects and doctors; Social Class II includes intermediate workers, such as shopkeepers, farmers and teachers; Social Class III Non Manual includes skilled non-manual workers, such as shop assistants and clerical workers in offices.

<sup>4</sup>1=under £35 pw; 2=£35-49 pw; 3=£50-99 pw; 4=£100-149 pw; 5=£150-199 pw; 6=£200-249 pw; 7=£250 or more pw.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika* 41(3), 321–327.