The causal effect of education on depression

ARNAUD CHEVALIER[♣]

and

LEON FEINSTEIN[♦]

October 2004
Preliminary draft – not for quotation

Abstract:
Low achievers tend to be more depressed individuals. This correlation may be the consequences of childhood characteristics and some unobservable characteristics explaining both educational attainment and mental health. We rely on a rich data set that allows us to control for depression at a young age and several estimation strategies to identify the causal effect of educational attainment on adult depression. At all level of education, women are more likely to be depressed than men and with the exception of university graduates, education appears to reduce the probability of depression. When accounting for possible bias due to omitted variables, the impact of education is strengthened and independent of additional controls. Propensity score matching estimates also are up to 20% larger. The effect of education persists as the individual age but we obtained conflicting conclusions on whether education affects the transition to depression.

Key words: Returns to education, health
JEL: I12, I29

Acknowledgment:
The data was made available by the Data Archive at Essex. The project was partially funded by the Evidence Based Policy Fund from Her Majesty Treasury. The authors thank participant at the LSE LoWER meeting (April 2004) as well as the late Mark Berger for comments. All remaining errors are solely ours.

[♣] Arnaud Chevalier, Department of Economics, University of Kent, and Centre for the Economics of Education, London School of Economics and the Institute for the Study of Labour, IZA, Bonn. Corresponding address: Department of Economics, University of Kent, Dublin 4, Canterbury, CT2 7NP,. a.chevalier@kent.ac.uk
[♦] Leon Feinstein, Institute of Education, Centre for Research on the Wider Benefits of Learning, 20 Berford Way, London WC1H 0AL, lf@bg.ioe.ac.uk

Introduction

Mental illness is a major driver of national health expenditure in the UK, a cause of substantial lost output in terms of sickness absence and non-employment and a source of considerable human cost in terms of reduced quality of life. Accounting for the first two sources, Patel and Knapp (1998) estimate that the total costs of mental illness in the UK were £32.1 billion in 1996/7. Kind et al. (1993) estimate that cost of depression stemming from lost productivity and increased benefit payments alone was £3 billions in Britain in 1993. Of this, the cost to the National Health Services was estimated at £420 million annually[1]. Recent estimates (Firte, 1999) suggest that the total cost may have risen to between £4 billion and £8 billion. Estimates for the US put the increased health costs for depressed individuals at $1500 (Henk et al., 1996). However, these costs exclude important personal costs such as suicide, strongly linked to mental illness (Mortensen et al., 2000), and costs for distress borne by families through reduced productivity. In fact, the World Health Organisation rated depression as one of the top three causes of disability and morbidity in the developed world (WHO, 1999). There is also evidence that the prescription of antidepressant drugs has been raising considerably (more than doubled) in the last 10 years.

Because of these large costs, partially borne by society, policies reducing the risk of depression may have large social returns. In this paper we focus on education as a potential policy affecting depression. There are a number of reasons why education and other learning interventions may have an effect on health (see Grossman, 1972, 2000 for extensive discussions). Education can directly affect heath outcome by making individual more able to process information and thereafter health conscious (Kenkel, 1991) or by improving the efficiency of treatment. For example, more educated are more prompt to seek diagnosis and more diligent in following treatment (Goldman and Lakdawalla, 2001 or Goldman and Smith, 2002). The effect of education

[1] The figure was subsequently quoted in The UK Department of Health White Paper, 'Saving Lives—Our Healthier Nation. 1999.' Anita Patel and Martin Knapp (1998) of the Centre for the Economics of Mental Health (Institute of Psychiatry and University of Kent), include this in their estimate of the total cost of mental illness in the UK of £32 million. Of this total the major items were £12 billion due to lost employment and productivity, £8 billion to DSS payments and £4 billion to NHS costs.

on health may also be indirect. Here, we hypothesise three main channels: (1) economic factors, i.e. income, employment or working conditions, (2) family relations, and, (3) psychosocial factors.

Estimating the causal effect of education on depression is an important policy issue, not only as a way to reduce the social cost of depression but also to estimates the returns to education. Adding the effects of schooling on mental health may substantially affect the returns, social and private to this investment[2]. Although it initially appears that more educated individuals are less likely to suffer from depression, we can only conclude that education reduces the risk of adult depression if the effect is causal. A negative correlation could merely indicate that some individual characteristics increasing the risk of depression are negatively related with educational achievement; these characteristics may be observable or non-observables, possible variables are cognitive ability, time preference or peer effects.

In this paper, we propose to identify the causal effect of education on depression. To do so, we rely on two distinct identification strategies: namely instrumental variables and propensity score matching. We also provide some tentative evidence on the channel by which education affects mental health but, we are mostly interested in the overall effect of education. We then estimate the monetary value of a policy increasing educational attainment. This estimated benefit represents a conservative estimate of the overall effect on the life course of the individuals concerned and omits intergenerational effects and other externalities. The outline of the paper is the following. The first section provides a rapid overview of the literature. This is followed by a discussion on the various identifying strategies. The British data used for the empirical work is described in section 3 and the results are discussed in section 4. The next section compares the results over the life time of the individuals and we conclude with some additional comments.

[2] See Haveman and Wolfe (1984) for a review of the different outcomes affected by education.

1 Literature

There is little evidence on the effect of education on depression. The concern of most epidemiological and etiological research in the field has been with the influences of genes (e.g. Zubenko et al., 2003), gene-environment interactions (e.g. Silberg et al., 2001), personality style (e.g. Kendler et al., 1993), prior history (e.g. Lewinsohn et al., 1988), adverse life events (e.g. Mazure et al., 2000), age and gender. Relative to these important features it is likely that education may operate only at the margin. Propper et al. (2004) using the BHPS note that more qualified individuals are significantly less at risk of bad mental health outcomes. Similarly, Kubzansky and Sparrow (1999) found that Americans with less than high school education were almost twice as likely to suffer from long-term stress as individuals with at least a college education even after controlling for age and lifestyle (smoking, alcohol consumption and exercise).

Education has been hypothesised as an important influence on psychosocial characteristics such as efficacy and self-esteem both of which have been found to have a moderator effect on depression. Seeman and Lewis (1995) found that adults who reported greater health problems tended to be those who had reported that they felt powerless five years earlier, controlling for their health at that time. Kennedy et al. (1988) found that exam stress was associated with weakened immunity.

Education may also influence life circumstances through its impact on labour market and/or marriage market prospects that may in turn provide the conditions for onset of depression. For example low income, unemployment, poor housing, unsatisfying or unstable relationships are all associated with loss of autonomy, vulnerability and stress, important possible pre-conditions for depression (Clark et al., 1992; Hammen et al., 1989; Segal et al., 1992) and have been found to impact on well-being (Blanchflower and Oswald, 2004) or Hauck and Rice (2003). The effect of occupation is ambiguous. Within a closely defined occupation such as White Hall civil servants, higher occupational grades were associated with more control over the working lives, more varied and challenging work and greater job satisfaction. These outcomes are associated with lower

hostility levels, fewer reported difficult life events, more healthy lifestyles, and lower rates of morbidity (Marmot et al., 1991). On the contrary (Rose, 2000) found that professionals and managers enjoy higher job satisfaction but also with higher levels of stress, whilst skilled workers and assembly-line workers experience relatively low levels of job-related stress. There may be important trade-offs between stress and satisfaction that may lead to a complex and non-linear relationship of educational success and mental health (Hartog and Oosterbeek, 1998).

This literature has provided measures of the correlation between education and mental health but has not been concerned in estimating causality. Indeed the causality could be reversed; Currie and Stabile (2004) or Heckman et al. (2004) provide evidence of the impact of non-cognitive skills such as attention, self-esteem and locus of control on educational attainment. The causality of education on other health outcomes is more established. Berger and Leigh (1989) pioneered the use of instrumental variable in health economics. Some of their exclusion variables relying on family characteristics may have been questionable but the most recent evidence have used educational reforms to identify the causal effect of education on birth weight (Currie and Moretti, 2004), mortality (Lleras-Muney, 2001) and smoking behaviour (Kenkel *et al.*, 2004).

To summarise, we posit three channels through which educational failure and failure to progress may create risk of depression, namely through effects on (i) psychosocial characteristics, in particular self-efficacy, (ii) the probability of negative working circumstances in terms of working conditions and work-satisfaction or negative life events in the field of labour market activity such as unemployment, and, (iii) similar probabilities of negative life circumstances or life events in the domain of domestic life, i.e. low satisfaction and stress in relationships with partners or children. Additionally, the effect of education on depression may be direct with, for example, more educated individual being more informed about various risk factors, more likely to seek treatment and more likely to follow the treatment. Whilst we provide some evidence by which

education may affect depression we are mostly interested in the overall impact of education on mental health.

2 The identification problem

Estimating the effect of education on adult depression is a typical problem of program evaluation; see Heckman, Lalonde and Smith (1999) for a general overview of this problem or Smith (1999) for a detailed review of the difficulties in estimating causal relationship between health and wealth. Here the treatment is defined as the level of education. It can be seen as a multiple treatment, varying in intensity (years of education) or form (qualification). To simplify the presentation, and the empirical analysis, we will concentrate on the simpler case of a single treatment, which in the following section will be defined as having at least O-levels as the highest qualification. In this section, we present alternative strategies to estimate the causal effect of education on depression.

We introduce the following notation: Y_1 and Y_0 are the depression status associated with the higher and lower level of education respectively. D is a dichotomous variable taking the value one for treated individual, those with more education. The parameter of interest is the average treatment effect on the treated (ATET).

$$ATET = E(Y_1 - Y_0 / X, D = 1) = E(Y_1 / X, D = 1) - E(Y_0 / X, D = 1) \quad (1)$$

The estimation problem is that individuals are observed in only one state of the treatment and the counter factual outcome $E(Y_0 / X, D = 1)$ is never observed. Assuming that individuals are randomly allocated to a treatment group, we have:

$$E(Y_0 / X, D = 1) = E(Y_0 / X, D = 0). \quad (2)$$

However, unless relying on an experimental set-up, it is not possible to assume that there is no self-selection and (2) generally does not hold. Heckman et al. (1998) decompose the self-selection bias in three components: B_1 is the bias that occurs due to lack of common support, B_2 arises from different distributions of X within the two populations on the common support, and B_3

is due to differences in outcomes that remain even after conditioning on observables and making comparisons on a region of common support (due to selection on unobservable characteristics). Since experimental data are impossible to obtain in most evaluations, econometricians have relied on various identifying assumption to recover the value of the counter factual outcome.

2.1 Homogeneous return model

Assume the following simple model of the determinants of depression for individual i (Y_i), where S represents the level of education, X_1 is a set of individual characteristics from periods concomitant or pre-dating the schooling decision, whilst X_2 is a set of individual characteristics posterior to the schooling decision that may be dependent on S.

$$Y_i = \beta_0 + \beta_1 S_i + \beta_2 X_{1i} + \beta_3 X_{2i} + \varepsilon_i \qquad (3)$$

The parameter of interest is β_1. If X_2 is excluded β_1 represents the total schooling effect. When X_2 is included, β_1 estimates the direct impact of education on depression after accounting for the potential channels of influence. Both estimates are of interest to the policy maker. The overall impact of education on depression measures the potential social returns to education whilst identifying the sources of the educational impact is crucial to design policies reducing the risk of depression[3]. Since the outcome of interest is dichotomous, (3) can be estimated by a probit model, in which we assume that the error terms are iid and normally distributed. This simple model suffers potentially from biases B_1 and B_3 and thus does not provide an estimate of the causal effect of education on depression.

2.2 Propensity score matching

The second identification strategy is to assume that we observe the determinants of the schooling decisions. We then match each individual i with a level of schooling S, with a control individual, similar to i on all observable characteristics but with a lower level of education. Rather

[3] However, the estimated coefficients β_3 will be biased since the variables X_2 are potentially correlated with unobservable characteristics explaining the choice of education and the health outcome of interest.

than match on all observable characteristics, it is equivalent to compute for each individual a likelihood (propensity) of reaching a given level of schooling (Pr (D=1/X)=P(X)) and match on this propensity (Rosenbaum and Rubin, 1983). The idea behind propensity score matching is that conditional on their propensity score (P(X)), matched individuals are identical, so that the counter-factual of the treatment for individual i is equal to the observed outcome for the matched control individual.

$$E(y_0 / X, D) = E(y_0 / X)$$
$$E(y_1 / X, D) = E(y_1 / X) \quad (4)$$

The treatment effect for the individual i is then simply the differences between the observed outcomes for i and the individual matched with i, and the average treatment effect on the treated is simply the mean of these differences. The crucial hypothesis in the implementation of propensity score matching is the Conditional Independence Assumption (CIA), which is equivalent to assuming that the selection into the treatment is only based on observable characteristics.

$$(Y_1 - Y_0) \perp D / X \quad \text{(CIA)}$$

Various methods exist to implement matching estimates, all based on the same strategy of pairing individuals but with different techniques for pairing or different weights given to counterfactual individuals. The generic formula is thus:

$$\hat{E}(Y_0 / \hat{P}(X_i)) = \sum_{j=1}^{J} w(\hat{P}(X_i), \hat{P}(X_j)) Y_{oj}$$

The expected depression with no treatment for an individual *i* with more education is a weighted average of the depression for individuals with less education. The weight being attached to a given control individual (*j*) is a function of individuals *i* and *j* propensity scores. Individuals who are the most similar to *i* in terms of observable characteristics (*X*) are given the largest weight[4].

[4] In the simplest matching method, the individual (*j*) whose propensity score is the closest to $\hat{P}(X_i)$ is given a weight of 1, all other individuals, receive a weighting of 0; this is known as nearest neighbour matching. Alternatively, all control observations can be given a weighting inversely related to their distance from $\hat{P}(X_i)$. The weighting is a kernel function, here we rely on Epanechnikov kernel. Matching can be conducted with or without replacement of the

If no match is found within the interval $[\hat{P}(X_i) - c_n, \hat{P}(X_i) + c_n]$, observation *i* is discarded. Discarded individuals indicate that the existence of common support is not a reasonable assumption. Thus, propensity score matching highlights the potential for bias B_1 to occur. Tests of the extent of common support provided by matching estimation can be used to assess the extent of bias B_1. In the absence of common support, probit and matching estimates will be biased.Whilst both matching and probit estimates assume selection in the treatment on observable characteristics, the matching estimate has two advantages, it highlights potential lack of common support between the treated and the non-treated, and it does not impose a functional form, therefore does not assume homogeneity of the treatment effect.

2.3 Instrumental variables

β_1 may be biased if omitted characteristics affecting mental health are also correlated with educational choice. This is equivalent to the rejection of the CIA assumption in favour of the alternative assumption that the selection into education is based on some unobservable characteristics of the individuals which may also be correlated with the outcome of interest. In order to get consistent estimates of β_1, it is possible to rely on instrumental variables. Instruments must be correlated with the decision to invest in education but be independent of subsequent mental health. We discuss our choice of instruments in the data section. This identifying strategy estimates eliminate the B_3 bias. We report the three types of estimates to highlight the relative magnitudes and directions of the three types of biases.

3 Data and methods adopted

The empirical analysis is conducted using the National Child Development Study (NCDS). The NCDS is a longitudinal study of all the children in Great Britain born in a given week of March 1958. These children and their parents were followed up at age 7, 11, 16. These questionnaires also included medical and school questionnaires. Members of the NCDS cohorts as well as their families were followed up at age 23, 33 and 42. These adult questionnaires also

control observations. Matching with replacement reduces bias but increase variance; due to the small sample size of the control group, this application uses matching with replacement.

included a health supplement. The outcome of interests is the malaise score, a large sample survey instrument designed to identify depression in non-clinical settings (Rutter et al., 1970). The score is calculated from responses to 24 questions on various symptoms of depression, mainly aspects of well-being or somatic condition. Individuals with more than 7 positive symptoms are coded as depressed (see Annex 1 for details). The scale tends to over-predict clinical depression (Meltzer et al., 1995). Depression is measured at age 23, 33, and 42 (Figure 1). At all three points, men are about 4 percentage points less depressed than women. The most noticeable pattern is the large increase in the proportion of depressed individuals between the age of 33 and 42, with the probability increasing from 8% to 16% for women and 4% to 12% for men. The correlation in depression between any two period ranges from .29 to .36; no gender differences are observed in these correlations.

The relationship of interest in this paper, between education and depression, is represented in Figure 2, for individuals aged 42. Similar trends are observed when depression is measured at age 23 or 33. Figure 2 reports the level of depression of individuals in terms of highest qualification obtained by age 33[5]. Overall, 12% of men and 16% of women are depressed[6]. As expected, there are large variations in the level of depression by education level, with more educated individuals less likely to be depressed. Individuals with no qualifications have the highest probability of depression; 25% of women with no qualifications are depressed. Each qualification from CSE to A-levels leads to a similar reduction in the risk of depression. The depression gap between female with no qualification and those with higher education reaches 17 percentage points; the proportion of depressed women with a degree is only a third of that observed for women with no qualifications. For all levels of education, men are less likely to be depressed than women. However, for men the effect of education on depression is not monotonous. Higher education has

[5] Since most education is obtained before the age of 23, the results do not differ when using highest qualification measured at age 23.
[6] In the UK, 7% of patients consult their doctor for depression (DoH, 1995). This statistics probably under report the extent of depression in the population.

an ambiguous effect on depression. For women, it reduces the probability of depression by 2 percentage points, but for men, it increases it by 3 percentage points. The increase in depression associated with the highest level of education may be an indication of the job related stress involved in occupations requiring a degree[7]. With the exception of university graduates, the gender gap is almost constant between educational levels; women are 5 percentage points more depressed than men. For individuals with a higher education qualification, the gap is reduced to 1 percentage point.

Before moving on to the main focus of this paper, it is of interest to check the relationship between depression and two measures of productivity: participation to the labour force and wages. These estimates are not meant to look for a causal effect of depression but simply stress the correlations between mental health and labour outcomes, in order to assess some of the personal costs faced by individuals suffering from depression. The estimates on the effect of depression are reported in Table 2. Depressed individuals are significantly less likely to be participating in the labour force; there is also some evidence that this correlation increases with age. Surprisingly, the estimate of the impact of depression on wages are all insignificant, which could indicate that our estimate of depression only captures the most severe cases of depression.

The omission in (3) of non observable factors correlated both with education and depression lead to a biased estimate of β_1. To simplify, we assume that some of these covariates, like genetic factors for example, are constant over time for individual i, whilst others, such as reflecting the current environment are time dependent. Since the focus is on adult depression, there is no simultaneity in the timing of both outcomes and the bias can only come from time independent factors as we exploit the richness of the data set and include measures of depression at an early age to capture the individual fixed characteristics explaining both educational choice and

[7] Hartog and Oosterbeek (1998) also report an inverted U-shape in satisfaction, with individuals holding mid-level qualification being the most satisfied.

depression. This strategy thus eliminates from B_3 type bias, the component due to fixed unobservable characteristics.

Thus, we report measures of depression at age 11 and 16 as well as other family characteristics during childhood that may affect adult depression as well as educational choice. Statistics on these control variables are summarised separately by gender in Table 1. Women are less likely to have no qualification than men (17% vs. 20%) but men are 4 percentage points more likely to have a higher education. Externalising and internalising behaviour are indicators of children's dominant behavioural responses to unobserved emotional disturbance (Rutter, 1967); externalisers acting up or behaving in a way linked to conduct disorder, internalisers being less naughty but more prone to depression (Feinstein et al., forthcoming). These behaviours are measured at age 11 on the parental questionnaire. At age 16, we rely on parental assessment of the behaviour of their child, to assess the mental state of the child. This assessment is based on 18 questions from the Rutter scale such as does the child appears irritable, miserable, destroy things…and the intensity of this behaviour (from 1 to 3).

Peer effects and more generally the neighbourhood in which one grows up may have long-term effects[8]. To capture some of this effect, we include the type of school attended at age 16, the proportion of pupils, broken down by gender, in the school attended by the child, who remained in post-compulsory education. The remaining control variables included are the child's ability in mathematics and English at age 7, since low ability at a young age may be correlated with psychological disorder (Beck, 1987), and family characteristics from birth until the age of 16. These include parental age at the birth of the child and their education, their interest in the child's education as perceived by the child's teacher, and the paternal social class. We include a dummy for whether the family experienced financial difficulties when the child was age 11 and 16. At age 16, we also control for the relationship between the child and the parent. First, we use an indicator

[8] Evidence on neighbourhood effects on depression only concern adults; Propper et al. (2004) using the BHPS and relying on a precise measure of neighbourhood (500 individuals) do not find systematic effect of neighbourhood characteristics (Disadvantage, mobility, age, ethnicity and urban-ness) on mental health or mental health transitions.

of whether the child lives with parental figures, but also the quality of these relationships in the view of the child. Whilst 73% of children get on well with their mother, the proportion on getting on well with father is only 67% for boys and 63% for daughters. The proportions of same sex school peer group staying on are identical between the two groups. Finally, boys in this 1958 cohort perform better than girls at age 7 in Mathematics but not in reading. On all other characteristics, there is no significant difference by gender.

To identify the exogenous effect of education on adult depression, instruments should be correlated with the education decision but uncorrelated with unobserved characteristics determining adult depression. We rely on two sets of identifying variables. First, we include teacher's expectations concerning the schooling of the child. To be more precise, this is the answer to the following question "in your opinion would staying-on past compulsory schooling benefit the child?" The second set of instrument is a series of statistics on a policy reform that affected some schools during the period. The British schooling system is divided between a public and private sector (Public schools), for the cohort of interest the latter catered for around 10% of pupils. Up to the Sixties, the public sector was divided into an academic track (grammar school) and a more vocational track (secondary modern schools). Pupils were tested at age 11 and depending on results were assigned to a track; switching track was not permitted. Grammar schools prepared pupils for the O-levels (taken at age 16) and A-levels (taken at age 18), both exams being necessary to enter higher education. During the Sixties, a reform of the public sector was introduced which stopped tracking and created schools providing equal opportunity for all children (Comprehensive schools). Whilst the change to comprehensive schooling was decided by the government, its implementation was left to the Local Education Authorities (LEA). Depending on their political, financial and demographic constraints LEA reformed their schools[9] either by merging schools from different tracks, changing the status of existing schools or creating new comprehensive schools.

[9] To these days, some LEAs still rely on tracking, or a mix of tracking and comprehensive schools

Thus independently of the initial child's ability, the quality of the teaching received in a comprehensive school varied depending on the conditions under which the school was formed. Pupils in a comprehensive schools formed out of a secondary modern are less likely to attend post-compulsory education than pupils in an amalgamated school and even less than those from a former grammar school (see also Harmon and Walker, 2000, for further discussion).

Table 3 reports the proportion of pupils achieving O-level and above by school type and for pupils attending a comprehensive school, by the origin of the school. As expected, pupils from the academic tracks are the most likely to have obtained O-levels or equivalent qualifications. More importantly, within comprehensive schools, there are significant differences in attainment of pupils depending on the origin of the school. Comprehensive that started as Grammar schools are more successful than others, and those that originated from a secondary schools are the least successful. This relationship still holds even after controlling for the initial ability of the pupil.

Statistics on the instruments are also presented in Table 1. Teachers are marginally more likely to think that girls would benefit from education more than boys, but overall, there is no gender differences in the mean values of the instruments. Since, the instruments are correlated with the decision to remain in post compulsory schooling, we recode our initial measure of education as a single treatment: having obtained O-levels or above rather than the 5 highest education categories previously mentioned.

To calculate the propensity scores we use all the variables from our initial exogenous model. Propensity scores are calculated for moving from one highest qualification to the next and also for the dichotomous treatment of having at least O-levels.

In additional models, we also include a vector of current characteristics (X_{it}). This vector of current characteristics (at the age when depression is measured) is broken down into three components: job, family and efficacy. Current job characteristics include the following: wage (in logarithm form), labour market experience since 16, whether currently working, whether works

full-time or part-time and a measure of self-employment. The family component reports marital status, whether had any children, number of children, age at first birth and age of the youngest child. Efficacy is introduced to assess the extent to which the channel for education effects is the sense of empowerment and control over life (Bandura, 1994) that has been theorised as a benefit of education (Hammond, 2002). A measure of global efficacy is available at ages 33 and 42, sample members asked to answer a set of questions such as "I never really seem to get what I want out of life," or "I usually get what I want out of life." The score is calculated as the sum of the answers to these questions.

Careful interpretation must be made of result with these three sets of measures included as they will interact strongly with the dependent variable depression and can be considered endogenous as they are probably affected by the same unobservable characteristics correlating education and depression. The base specification provides an estimate of the total effect of education on depression whilst these subsequent specifications are intended only to indicate the process by which education affects depression. These estimation strategies are used when the outcome of interest is depression at age 23, age 33 or age 42. Full results are only reported for the oldest age, but Tables summarising the evolution of the effect of education over the life time are also presented.

4 Results

Our goal is to identify the causal effect of education on depression in (3). Rather than attempting to estimate complex structural models that describe in detail the channels by which education impacts on depression, which despite the richness of the data does not seem feasible, we rely on a reduced form model. For this approach to be satisfactory, we need to exclude covariates that are affected by the educational choice made. Thus, our base specification does not include any adult characteristics (i.e. $X_2=0$).

4.1 The basic model

Table 4 reports the marginal effects of covariates on the probability of depression at age 42. The first two columns estimate, separately by gender, the base model when education is measured by the highest qualification obtained by age 33 and education is assumed to be exogenous.

Measures of mental health at younger age significantly affect adult depression. For women, an additional point on the internalising score increases the depression probability at age 42 by 1.4 percentage points, whilst for men, externalising score at age 11 is relevant. There is persistence in propensity to depression over the life-time which is consistent with the assumption that some of the unobserved characteristics leading to depression are time-independent. Similarly, the parental evaluation of the character of teenagers at age 16 predicts depression. On this score, ranging from 17 to 48, each additional point is associated with an increase in the probability of depression by 0.5 percentage points.

Whether teenagers live in an intact family does not affect their adult depression score but the quality of the relationship with the parents does matter. The effect is stronger for the parent of the opposite sex; for both genders, getting on well with parents as a teenager reduces adult depression by four to five percentage points. Finally, financial hardship significantly increases the probability of adult depression but only for women. Women experiencing financial difficulties at age 16 are 8 percentage points more at risk of depression. This effect is similar in magnitude to gaining a qualification and could be due to persistence in income over generation.

As in the statistical analysis, education reduces the risk of depression for both genders; but for men the effect is statistically significant only for individuals with at least 0-levels. For women, having a CSE reduces the risk of depression by three percentage points, any additional qualification up to A-level reduces this probability by a further two percentage points. Higher education does not lead to a significant further reduction in depression compared to A-levels. For men, the pattern is different. Having a CSE does not lead to a significant reduction in adult depression, but reaching O-

levels leads, as for women, to a reduction of about five percentage points. For men, the effect of education is non-linear, and whilst A-level leads to a further reduction in the probability of depression, individuals with higher education as their highest qualification are at the same level of risk as individuals with O-levels. Getting on well with the parent of the opposite sex, has the same effect on reducing depression as having CSEs relative to no qualifications. For women, not having experienced financial hardship at age 16, is equivalent to having gained A-levels rather than no-qualification. So while the effects of education on depression are significant, family characteristics are at least as important.

As explained in the previous section, the education coefficient may be biased if some unobserved characteristics are correlated both with education and depression. Since the correlation between education and depression is negative, the coefficient on schooling would be biased upwards[10]. To account for the possible endogeneity of education, we estimate the basic model using instrumental variables. As argued above, our instruments do not allow us to differentiate between all qualifications therefore we redefine the measure of education as having at least O-levels. For comparison purposes, we first estimate a model, where we assume the exogeneity of this educational choice; this is reported in Column 3 and 4 of Table 4. The control variables have almost identical effect as in the multi-level qualification models; thus we do not comment them. Having at least O-levels reduces depression for both men and women by about 5 percentage points. Allowing for the endogeneity of education, the reduction in depression due to education is further strengthened. This is a surprising result which is not likely to be driven by a Local Average Treatment Effect unless individuals affected by the school reform were more likely to be depressed. The downward bias of the probit estimates may stem from measurement error in education. Alternatively, the estimates suggest that individuals investing in education post-16 would be at greater risk of depression in the absence of further or higher education; for example individuals

[10] In regressions excluding past depression measures (not reported here), the estimated effects of qualification was between 0.5 and 2 percentage points higher than in the reported model.

with a greater ability or taste for analysing problems are more likely to invest in education but may also be more self-critical which leads to a greater risk of depression. Empirically, more educated individuals are found to be less happy (Hartog and Oosterbeek, 1998). Moreover, Feinstein and Banner (2004) find that children with greater internalising behaviour are more likely to stay on in school and more prone to depression as adults since internalising behaviour is associated with low self-esteem and poor peer relations

The instruments are significant in the first stage equation, and pass the rule of thumb for weak instruments[11]. The over-identification also shows that the instruments are valid. For men, but not for women, we reject the exogeneity test, so the IV estimates are only significantly different from the probit estimates for women.

4.2 Routes through which education may affect depression

The effect of education identified in this reduced form is the total effect and has various sources. Education affects career path, family characteristics and non-cognitive skills, feeding back on mental health. We are not interested in the effect of income (or other post-education variables) on depression. Our intention is simply to take out the partial correlation between schooling and depression that stems from income. This may over-control since the other channels may also be correlated with income, but does indicate how much of the effect of education on depression does not pass through income. With income entered one can observe how much residual correlation remains that is orthogonal to the relationship of income, schooling and depression.

Results are reported in Table 5 for 4 types of model, by OLS and IV. Model 1 includes work characteristics. The inclusion of these variables reduces the effect of education on depression by 20% to 40% depending on gender and whether education is assumed exogenous. It can also be noted, that for men we cannot reject the hypothesis that the IV estimates are not significantly

[11] Bound et al. (1995) recommend that a F-test on the joint significance of the instruments in the first stage equation be greater than 10, in order not to suffer from weak instruments problems.

different from the estimate assuming exogeneity of education. This is also true for all subsequent models.

Model 2 controls for current family characteristics. The issue here is that education may alter fertility decisions (leading to a reduction in number of children, older age at first birth) and marital situation (assortative mating) which potentially affect the depression probabilities. However, the inclusion of these family characteristics does not have any significant impact on the estimated effect of education on depression. Thus it does not appear that the effect of education on depression is through the current family characteristics, though the true story may be a more complex one involving interactions between work characteristics and family characteristics.

In model 3, we add the measure of efficacy to the base specification. This has the expected effect of reducing the impact of education by about 25% for all specifications. Finally, in Model 4, we include all three sets of variables, work, family and efficacy. The estimates on education become less precise at around a reduction in risk of depression of 10 percentage points for women and men. It appears that the effect of education on depression is partly through a work effect (income effect, or more satisfying participation), partly though an improvement in efficacy, but that a large component is unexplained by the channels explored here and is a measure of the direct effect of education on depression.

4.3 Propensity score matching

Tables 5 and 6 report estimates of the effect of education on depression obtained by propensity score matching. Since several models rejected the endogeneity of education assumption, the assumption that the selection into education is solely based on observable characteristics may be acceptable. Table 6 replicates the first model presented in Table 4, and estimates the effect of having a given qualification rather than the one at the previous level. For each comparison, we estimate a probit model of the probability of having education S_j rather than S_{j-1}. The specification

includes all the covariates of the base model. Matching is reported using nearest neighbour with replacement and a calliper of 0.01, as well as Epanechnikov kernel and a bandwidth of 0.01[12].

Additionally, the propensity score matching provides an estimate of the average effect of the treatment on the treated only if all treated individuals are actually matched to a control. If this common support assumption is not fulfilled, the estimate is biased (Heckman et al., 1997). Thus, we report the proportion of treated individuals that are successfully matched. For all models, this proportion is at least 96%, so the bias due to the lack of common support may be small. Additionally, we provide information on the extent of the common support, by reporting the number of control observations responsible for 50% of the matches. In all cases a substantial number of controls are used, which gives us some confidence in the robustness of the results. Since we have a limited sample, estimates do vary according to the matching procedure used but are reasonably similar to give us some confidence in this estimation strategy. Relying on the kernel estimates, the main effect of education on depression is for women, to move from no qualification to CSE. This effect is twice as large as the one estimated parametrically. Any further qualification does not significantly affect the risk of depression. For men, the effects are imprecisely estimated and no qualification significantly affect the level of depression.

In Table 7, we report propensity score matching estimate in a model where education is measured by having at least O-levels. The distribution of propensity scores is reported in Figure 3 separately by gender. Whilst the distribution for controls and individuals with at least O-levels are different, there is common support throughout. Once again, we report results for two types of matching process but also for two calliper/bandwidths, which highlights the possible bias due to lack of common support for individuals with high propensity scores. With the tighter bandwidth between 93% and 98% of individuals are matched, while when using the larger calliper all individuals are matched to a control. The bias due to lack of common support appears to be small in

[12] Results with calliper/bandwidth set at 0.05 and 0.003 are not substantially different.

most cases; only for men when using the kernel match, do we find substantially different estimates according to the bandwidth used. Estimates are stable to the choice of matching process and attainment of O-levels reduces the risk of depression by 6%. This is about 20% larger than the estimates from the probit models, which also assume selection on the observables, and suggest that imposing a functional form leads to a large bias.

5 Depression over time

The results presented in Tables 8 and 9 assess the evolution of the education effect over time. More specifically, we focus on depression at age 23 and 33. Table 8 reports the results for models using the specifications presented in Tables 4 and 5, assuming exogeneity or endogeneity of education. We introduce contemporary controls for work, family and efficacy at each age.

In the base model, the effect of education on depression is remarkably persistent, especially for women. In models assuming the endogeneity of education, the estimates are about three times as large as in the exogenous models. Independently of additional covariates, having at least O-levels reduces the probability of depression at age 23 and 33, for about 13 percentage points for women. The channel for this reduction appears to be unexplained by our potential channels, and largely independent of income effects, changes in work, family or efficacy characteristics.

For men, we cannot reject the exogeneity of education at the 5% level in any models. Focusing on models assuming exogeneity, the estimated effect of education is similar between all specifications, so again the process by which education reduces depression remains unexplained. For men, the relationship between education and depression appears to be U-shaped with age. This is confirmed in Table 9, where estimates from propensity score matching are presented. For women, the pattern is not as clear and depends on the model specification and estimation technique.

Finally, we look at the effect of education on depression transitions. Keeping only individuals who were not depressed in the previous wave, we estimate the effect of education on the risk of becoming depressed. This is done for two transitions, at age 33 and age 42. These results

are ambiguous. For women, each additional qualification significantly reduces the risk of becoming depressed. The effects of each qualification are remarkably similar, suggesting again that the most important educational effect is found for moving individuals with no qualification to some qualification. When defining education as having O-levels or more, no significant effect is found either in the endogenous or exogenous models. For men, no effect of education on depression transitions is found.

6 Depression as a continuous measure

So far this paper has defined depression as a dichotomous variable. Whilst this may be desirable in the medical literature if the outcomes of interest are likelihood of using some type of treatment, it may underestimates the effect of mental health on labour outcomes, as individual with low malaise score may already have a reduced productivity. Possibly, the impact of education on mental health was also underestimated when using this unitary measure of depression, for example if education reduces the malaise score so that individuals do not cross the ad-hoc threshold. Thus, we also now present estimates of the impact of education on the malaise score. Looking at the distribution of malaise score in Appendix 1, it is easy to see that malaise score as a skewed distribution, with an over-representation of zeros and a thin long tail of individuals with high score. Since the score only takes 25 distinct values at most, it is possible to use a count data analysis accounting for the high propensities of zeros and over-dispersion. The estimates of a zero inflated negative binomial model are reported in Table 11 by gender and age. Specifications using highest qualification as a measure of education lead to similar conclusions to those presented above (need to calculate marginal effects) with each qualification reducing the malaise score. In the model using achieving O-levels or above as an exogenous measure of education, the coefficients are stable over the life time of the individuals and of similar magnitude between genders. Moving on to assuming the endogeneity of education, the estimates are much more precise than those obtained when a dichotomous measure of mental health was used. Furthermore, for women and men aged 42, the estimated coefficients are half as large as those obtained in the exogenous model. This is in contradiction with the models previously presented. Using all the information on mental health it appears that unobservable characteristics correlated with education achievement are negatively

correlated with mental health, however, policies increasing educations would still have a substantial effect at reducing malaise scores.

7 Conclusion

Using a large array of specifications, we consistently find that education significantly reduces the risks of adult depression. The effect is non-linear and is larger at low level of education. Estimates using two-stage least squares are much larger but in most cases, it was not possible to reject the exogeneity of education. Using propensity score matching, we estimate that individuals with at least O-levels reduce their risk of adult depression by 6 percentage points. This effect is similar for men and women and robust to the choice of calliper or matching process. Imposing a functional form biases the estimates downwards. The positive effect of education is present at all ages and remains even after accounting for work and family characteristics or efficacy. There are some mixed evidences that gaining some qualifications also reduces the risk of transition to depression.

These results suggest that policy increasing the education of individuals would have positive effects on their future mental health. It is not possible yet to identify where this effect stems from. It is difficult to calculate the cost benefit of such policies since the costs of depression, while assumed to be large have still not been satisfyingly identified and the costs of a policy increasing education can vary substantially. Machin and McNally (2004) for example identify that the literacy hour programme has an extremely low cost and significantly affects educational attainment whilst other policies such as increasing teachers' wages or reducing class size, are costly and their effects are still controversial.

Figure 1: Depression over life time

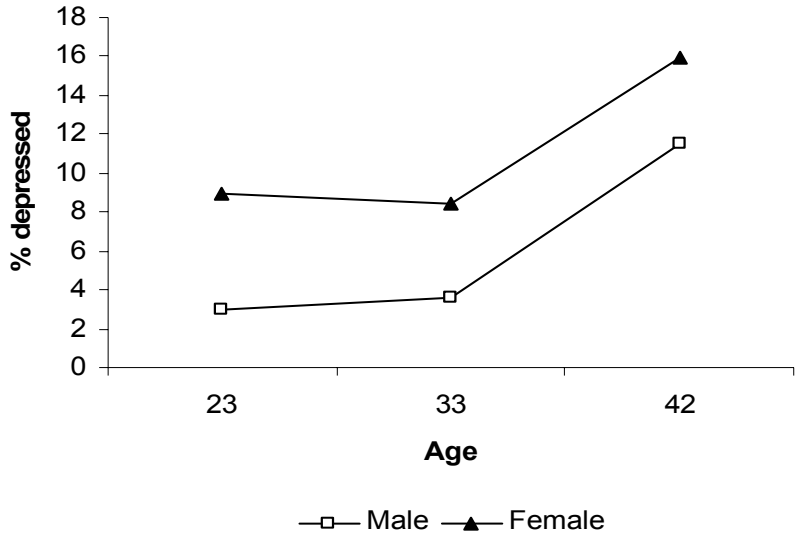


Figure 2: Education level and depression at age 42

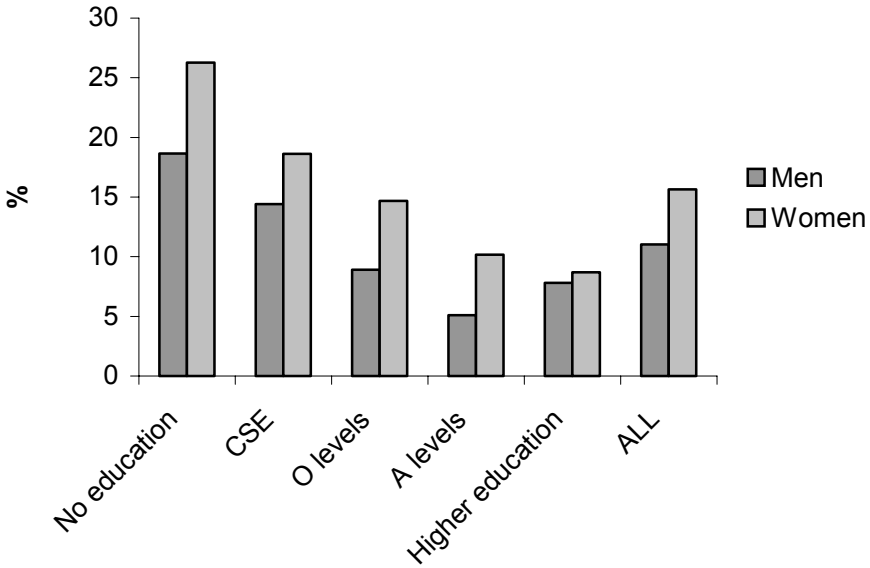


Figure 3: Depression and job satisfaction

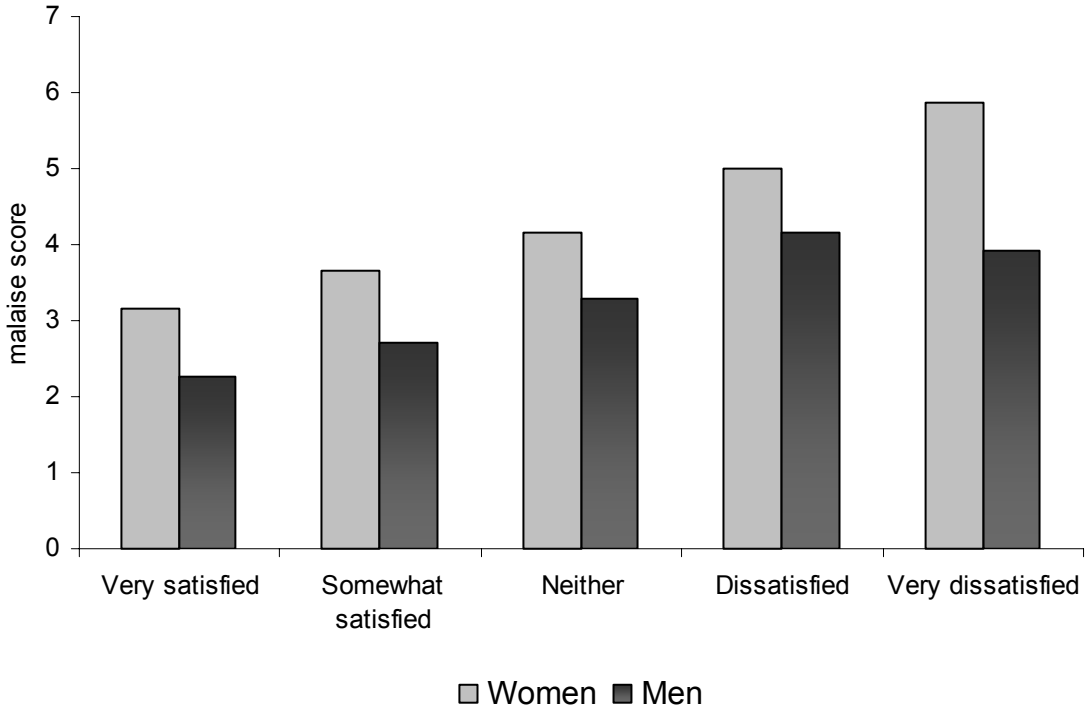
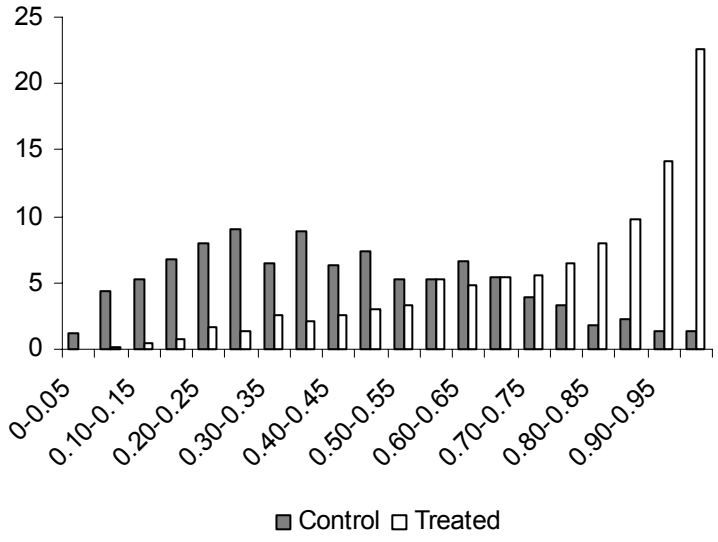


Figure 4: Distribution of propensity score (O level and above)

A] Men



B] Women

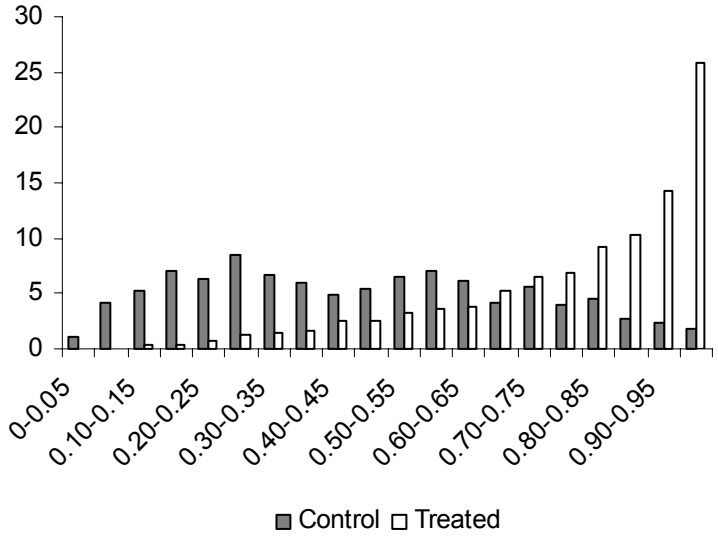


Table 1: Summary statistics of selected variables

	Men		Women	
Variable	Mean	Std. Dev.	Mean	Std. Dev.
Depression 42	0.115	0.319	0.160	0.367
Depression 33	0.030	0.169	0.090	0.286
Depression 23	0.035	0.185	0.083	0.277
CSE	0.166	0.372	0.141	0.348
O-level	0.377	0.485	0.447	0.497
A-level	0.083	0.275	0.096	0.295
Higher education	0.183	0.387	0.147	0.354
dad age when born	29.908	7.590	29.558	8.178
mom age when born	27.550	5.468	27.514	5.643
dad age left education	16.041	1.761	16.071	1.746
mom age left eduation	16.077	1.527	16.157	1.620
internal score @ 11	6.510	1.564	6.568	1.593
external score @ 11	9.019	2.132	8.409	1.987
financial difficulties @ 11	0.088	0.284	0.100	0.300
character @ 16	21.839	3.827	21.991	3.748
Get well with Mother	0.732	0.443	0.734	0.442
Get well with Father	0.673	0.469	0.637	0.481
Live with mother	0.955	0.207	0.957	0.204
Live with father	0.895	0.306	0.873	0.333
financial difficulties @ 16	0.084	0.277	0.093	0.291
% boy staying on	60.048	24.721	55.398	20.052
% girl staying on	55.005	20.668	60.143	24.884
Reading test 7	-0.082	1.011	0.172	0.881
Math test 7	0.059	0.978	-0.028	0.963
Type of school: Grammar	0.098	0.297	0.117	0.322
Type of school: Comprehensive				
Type of school: Secondary	0.184	0.389	0.195	0.396
Type of school: Independent	0.045	0.208	0.043	0.203
Type of school: Other and missing	0.152	0.337	0.086	0.203
Instruments				
Child would benefit from further schooling (teacher's view)	0.372	0.483	0.418	0.493
Grammar to comprehensive	0.100	0.300	0.104	0.306
Secondary to comprehensive	0.136	0.343	0.119	0.323
Amalgamate comprehensive	0.149	0.356	0.136	0.343
Observations	3271		3395	

Note: Observations for depression at age 23 and 33 are respectively for men and women: 2837, 2820 and 3033 and 3073

Table 2: Effect of depression on productivity

	Women		Men	
Panel A	Labour force participation	Ln. wage	Labour force participation	Ln. wage
At age 23	-0.100	-0.222	-0.001	0.211
	(0.039)	(0.186)	(0.030)	(0.287)
At age 33	-0.082	-0.045	-0.112	0.002
	(0.034)	(0.052)	(0.034)	(0.061)
At age 42	-0.178	-0.016	-0.154	-0.034
	(0.022)	(0.027)	(0.021)	(0.039)

Note: The specifications include highest qualification, test scores in Math and English at age 7, type of school. In the wage regression, we also control for experience and whether working part-time. In the participation equation, the additional controls are: marital status, number and age of the youngest child.

Table 3: School type and post-compulsory schooling

	Proportion with O-level and above	Observations
Type of school		
Secondary	54.35	1264
Comprehensive	64.86	3441
Grammar	94.84	717
Public school	91.52	295
Origin of comprehensive school		
Secondary	58.07	849
Grammar	71.55	682
Amalgamated	64.81	949
Other/ purposed built	66.45	921

Table 4: Determinants of depression at age 42- Marginal effects

	Probit		Probit		IV	
	Female	Male	Female	Male	Female	Male
Education						
CSE	-0.034	-0.008				
	(0.018)	(0.016)				
O levels	-0.050	-0.045				
	(0.017)	(0.014)				
A levels	-0.077	-0.067				
	(0.018)	(0.014)				
Higher education	-0.078	-0.048				
	(0.018)	(0.016)				
O-level or above			-0.043	-0.050	-0.211	-0.103
			(0.016)	(0.014)	(0.084)	(0.070)
Internal score @ 11	0.015	0.003	0.014	0.003	0.016	0.004
	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)
External score @ 11	-0.002	0.006	-0.001	0.006	-0.005	0.006
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)
Character score @16	0.007	0.004	0.007	0.004	0.007	0.004
	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)
Get well with dad @ 16	-0.050	-0.019	-0.051	-0.019	-0.053	-0.019
	(0.017)	(0.016)	(0.017)	(0.016)	(0.017)	(0.016)
Get well with mum @16	-0.002	-0.043	-0.003	-0.044	0.004	-0.041
	(0.018)	(0.019)	(0.018)	(0.019)	(0.018)	(0.019)
Financial hardship @16	0.078	0.019	0.080	0.020	0.099	0.017
	(0.025)	(0.020)	(0.025)	(0.020)	(0.026)	(0.021)
Observations	3395	3271	3395	3271	3395	3271
Pseudo R^2	0.080	0.079	0.078	0.077	0.078	0.071
Significance of instrument : $\chi^2(4)$, p					194.4 p=0.00	167.9 p=0.00
Exogeneity test: $\chi^2(1)$, p					5.29 P=0.02	0.665 P=0.41
Overidentification test[A]: $\chi^2(3)$, p					4.71 p=0.19	0.738 p=0.86

Note: Robust standard errors are reported into parentheses.

The basic regression also includes controls for parental age and education; parental interest in child's schooling at age 11, paternal social class at 11, financial hardship at age 11 and whether leaving with natural parents at age 16, as well as test score at age seven in math and English, the type of school attended at age 16 and the proportion of pupils at the school attended who stayed on after compulsory education, region of residence at age 16 and at age 42.

Instruments for O levels and above include: teacher's view on whether child would benefit from more schooling and for children attending a comprehensive schools, how this school was created.

[A] Overidentification test reports the Hansen J, distributed as a $\chi^2(3)$. This statistic is calculated using a linear model estimated by GMM. The estimates are similar to those presented.

Table 5: Determinants of depression at age 42 (including additional characteristics)- Marginal effects

	Model 1		Model 2		Model 3		Model 4	
	Female	Male	Female	Male	Female	Male	Female	Male
Education								
O-level and above	-0.026	-0.035	-0.039	-0.050	-0.028	-0.041	-0.016	-0.030
(probit)	(0.015)	(0.013)	(0.015)	(0.013)	(0.014)	(0.013)	(0.014)	(0.013)
O-level and above	-0.169	-0.104	-0.189	-0.109	-0.149	-0.091	-0.112	-0.102
(IV)	(0.086)	(0.073)	(0.083)	(0.073)	(0.080)	(0.068)	(0.082)	(0.075)
Work characteristics	Yes	Yes					Yes	Yes
Family			Yes	Yes			Yes	Yes
Characteristics								
Efficacy					Yes	Yes	Yes	Yes
Observations	3395	3271	3395	3271	3395	3271	3395	3271
Exogeneity test: $\chi^2(4)$	3.54	1.09	4.24	0.77	2.89	0.64	1.71	1.13
Probability	P=0.06	P=0.29	P=0.04	p=0.38	p=0.09	P=0.42	p=0.19	p=0.29

Note: Robust standard errors are reported into parentheses. The specification includes the same covariates as specified in Table 1.
Model 1 adds log pay, and months of work experience as well as dummies for not working, self-employment, full-time work and pay missing.
Model 2 adds a set of dummies for marital status, number of children, age at first child, age of the youngest child and a dummy for no child.
Model 3: adds efficacy score and a dummy if efficacy score is missing. (Efficacy is not reported at age 23).
Model 4 includes all the variables of the base model as well as the additional variables of Model 1,2 and 3.
Instruments for O levels and above include: teacher's view on whether child would benefit from more schooling and for children attending a comprehensive schools, how this school was created.

Table 6: Effect of qualification on depression at age 42 (propensity score matching)

	Propensity score matching			
	Female		Male	
*** CSE vs no qualification**				
Nearest neighbour match	-0.083	(0.046)	-0.022	(0.033)
Epanechnikov kernel	-0.061	(0.035)	-0.017	(0.024)
Observation / Treated	1052 / 477		1170 / 544	
% matched treated	99		99	
Nbr controls used for 50% match	21		81	
***O-levels vs CSE**				
Nearest neighbour match	-0.009	(0.031)	-0.037	(0.030)
Epanechnikov kernel	-0.021	(0.026)	-0.037	(0.026)
Observation / Treated	1994 / 1517		1776 / 1232	
% matched treated	99		99	
Nbr controls used for 50% match	76		80	
*** A-levels vs O-levels**				
Nearest neighbour match	-0.025	(0.034)	-0.038	(0.034)
Epanechnikov kernel	-0.029	(0.028)	-0.033	(0.024)
Observation / Treated	1844 / 327		1492 / 270	
% matched treated	98		97	
Nbr controls used for 50% match	102		76	
*** Higher education vs A-levels**				
Nearest neighbour match	-0.017	(0.033)	0.014	(0.027)
Epanechnikov kernel	-0.010	(0.026)	0.020	(0.020)
Observation / Treated	825 / 498		863 / 593	
% matched treated	96		98	
Nbr controls used for 50% match	46		46	

For nearest neighbour match the calliper was fixed at 0.01. For kernel match, the bandwidth was fixed at 0.01. Standard errors are obtained by bootstrap with 500 replications.

Variables included in the propensity score are:

Parental age and education; whether parents read to the child at age 11, paternal social class at 11, internal and external score at age 11, character score at age 16, getting on with parents at age 16, financial hardship at age 16, whether leaving with natural parents at age 16, the proportion of boys and girls in the school attended who stayed on in post-compulsory education, test score in Math and English at age 7.

Table 7: Effect of O-level and above on depression at age 42 (propensity score matching)

	Propensity score matching	
	Female	Male
Education: O levels or above		
Nearest neighbour		
Calliper = 0.01	-0.062	-0.061
	(0.032)	(0.032)
Observation / Treated	3394 / 2343	3270 / 2100
% matched treated	100	100
Nbr controls used for 50% match	53	47
Calliper = 0.003	-0.050	-0.056
	(0.025)	(0.024)
Observation / Treated	3394 / 2342	3270 / 2100
% matched treated	96	91
Nbr controls used for 50% match	55	58
Epanechnikov kernel		
Bandwidth = 0.01	-0.055	-0.061
	(0.028)	(0.029)
Bandwidth = 0.003	-0.044	-0.062
	(0.022)	(0.021)

The specification of the propensity score is identical to the one described in the note under Table 3. Standard errors are obtained by bootstrap with 500 replications.

Table 8: Determinants of depression over time (including additional characteristics)- Marginal effects

	Model 0		Model 1 (Work)		Model 2 (Family)		Model 3 (Efficacy)		Model 4 (All)	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
Depression at age 23										
O-level and above (probit)	-0.052	-0.024	-0.037	-0.019	-0.032	-0.019	NA	NA	-0.031	-0.017
	(0.012)	(0.007)	(0.012)	(0.007)	(0.012)	(0.007)			(0.012)	(0.007)
O-level and above (IV)	-0.130 **	-0.055	-0.012	-0.123 *	-0.003	-0.175 *	NA	NA	-0.006	-0.143
	(0.044)	(0.028)	(0.050)	(0.092)	(0.055)	(0.146)			(0.059)	(0.132)
Depression at age 33										
O-level and above (probit)	-0.044	-0.011	-0.039	-0.009	-0.037	-0.009	-0.030	-0.005	-0.025	-0.003
	(0.012)	(0.007)	(0.012)	(0.006)	(0.012)	(0.006)	(0.011)	(0.006)	(0.011)	(0.004)
O-level and above (IV)	-0.078	-0.052 *	-0.074	-0.057	-0.025	-0.036	-0.028	-0.039	0.000	-0.034
	(0.070)	(0.054)	(0.069)	(0.057)	(0.059)	(0.047)	(0.055)	(0.046)	(0.047)	(0.044)
Depression at age 42										
O-level and above (probit)	-0.043	-0.050	-0.026	-0.035	-0.039	-0.050	-0.028	-0.041	-0.016	-0.030
	(0.016)	(0.014)	(0.015)	(0.013)	(0.015)	(0.013)	(0.014)	(0.013)	(0.014)	(0.013)
O-level and above (IV)	-0.211 **	-0.103	-0.169 *	-0.104	-0.189 **	-0.109	-0.149 *	-0.091	-0.112	-0.102
	(0.084)	(0.070)	(0.086)	(0.073)	(0.046)	(0.073)	(0.080)	(0.068)	(0.082)	(0.075)

Note: Robust standard errors are reported into parentheses. The specification in Model 0 includes the same covariates as specified in Table 1.
Model 1 adds log pay, and months of work experience as well as dummies for not working, self-employment, full-time work and pay missing.
Model 2 adds a set of dummies for marital status, number of children, age at first child, age of the youngest child and a dummy for no child.
Model 3: adds efficacy score and a dummy if efficacy score is missing. (Efficacy is not reported at age 23).
Model 4 includes all the variables of the base model as well as the additional variables of Model 1, 2 and 3.
Instruments for O levels and above include: teacher's view on whether child would benefit from more schooling and for children attending a comprehensive schools, how this school was created.
*,** indicates 10% and 5% statistical significance in the exogeneity test.

Table 9: Effect of O-level and above on depression over time (propensity score matching)

	Depression		Depression		Depression	
	Age 23		Age 33		Age 42	
	Female	Male	Female	Male	Female	Male
Education: O levels or above						
Nearest neighbour						
Calliper = 0.003	-0.010	-0.024	-0.045	-0.014	-0.050	-0. 056
	(0.020)	(0.014)	(0.019)	(0.011)	(0.025)	(0.024)
Epanechnikov kernel						
Bandwidth = 0.003	-0.026	-0.021	-0.050	-0.005	-0.044	-0.062
	(0.017)	(0.013)	(0.017)	(0.010)	(0.022)	(0.021)

The specification of the propensity score is identical to the one described in the note under Table 3. Standard errors are obtained by bootstrap with 500 replications.

Table 10: Effect education on depression transitions

	Women			Men		
	Probit	Probit	IV	Probit	Probit	IV
Age 23 - 33						
CSE	-0.019			0.008		
	(0.009)			(0.009)		
O levels	-0.017			0.001		
	(0.009)			(0.007)		
A levels	-0.031			-0.009		
	(0.008)			(0.006)		
Higher education	-0.023			-0.005		
	(0.009)			(0.008)		
O-level or above		-0.011	-0.008		-0.005	-0.009
		(0.009)	(0.033)		(0.006)	(0.022)
Pseudo R^2	0.082	0.076	0.077	0.099	0.094	0.098
Age 33 - 43						
CSE	-0.037			0.006		
	(0.017)			(0.016)		
O levels	-0.036			-0.029		
	(0.017)			(0.013)		
A levels	-0.050			-0.037		
	(0.018)			(0.016)		
Higher education	-0.046			-0.024		
	(0.018)			(0.017)		
O-level or above		-0.018	-0.050		-0.036	-0.065
		(0.015)	(0.052)		(0.013)	(0.044)
Pseudo R^2	0.055	0.052	0.052	0.066	0.065	0.061

Note: keep only individuals who reported not being depressed in the first period

Table 11: Effect of education on malaise score

	Female			Male		
	zinb	zinb	IV	zinb	zinb	IV
Age 42						
CSE	-0.080			-0.168		
	(0.054)			(0.065)		
O-levels	-0.176			-0.229		
	(0.047)			(0.058)		
A-levels	-0.327			-0.354		
	(0.070)			(0.085)		
Higher ed.	-0.289			-0.356		
	(0.063)			(0.075)		
O-levels and above		-0.680	-0.361		-0.535	-0.266
		(0.166)	(0.131)		(0.149)	(0.168)
Age 33						
CSE	-0.116			-0.157		
	(0.073)			(0.082)		
O-levels	-0.313			-0.289		
	(0.063)			(0.076)		
A-levels	-0.397			-0.326		
	(0.093)			(0.108)		
Higher ed.	-0.470			-0.485		
	(0.089)			(0.102)		
O-levels and above		-0.778	-0.494		-0.442	-0.547
		(0.151)	(0.177)		(0.117)	(0.220)
Age 23						
CSE	-0.009			-0.221		
	(0.055)			(0.072)		
O-levels	-0.216			-0.364		
	(0.048)			(0.066)		
A-levels	-0.335			-0.370		
	(0.079)			(0.102)		
Higher ed.	-0.370			-0.478		
	(0.071)			(0.092)		
O-levels and above		-0.819	-0.546		-0.501	-0.780
		(0.139)	(0.141)		(0.107)	(0.198)

Note: see note under Table 4 for specification. Standard errors in the IV models are obtained by boostrapping (500 replications)

References:

Bandura, A. (1994), "*Self-efficacy: The exercise of control*" New York: W. H. Freeman.

Blanchflower D. and A. Oswald (2004) "Well-being over time in Britain and the USA", *Journal of Public Economics*, 88, 1359-1386

Berger M. and P. Leigh (1989), "Schooling, self-selection and health", *Journal of Human Resources*, 24, 433-455

Bound J., D. Jaeger and R. Baker, 1995, "Problems with instrumental variables estimation when the correlation between the instruments and the endogeneous explanatory variable is weak", *Journal of the American Statistical Association*, 90, 443-450

Currie J. and E. Moretti, (2003), Mother's education and the intergenerational transmission of human capital. Evidence from college openings and longitudinal data, *Quarterly Journal of Economics*, 118, 1495-1532

Currie J. and M. Stabile (2004) "Child mental health and human capital accumulation: the case of ADHD", *National Bureau of Economic Research*, WP 10435

DoH (1995) The health of the Nation: Key are handbook – mental illness. 2nd Edition, HMSO

DoH (1999) Services for people who have depression, Clinical Standards Advisory Group, HMSO

Feinstein, L. and Bynner, J. (Forthcoming 2005) "Continuity and discontinuity in mid-childhood: implications for adult outcomes in the UK 1970 Birth Cohort," in Huston, A. and Ripke, M. (Eds) "Middle Childhood: Contexts of Development." Harvard University Press, Cambridge Mass..

Goldman D. and D. Lakdawalla (2001). Understanding health disparities across education groups, *National Bureau of Economic Research*, WP 8326

Goldman D. and J. Smith, 2002, Can patient self-management help explain the SES health gradient?, *Proceedings, National Academy of Sciences*, 99, 10929-34

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

Grossman M. (2000), The human capital model, in the *Handbook of health economics*, A. Cuyler and P. Newhouse (eds), North Holland

Hammond, C. (2002). *Learning to be Healthy*. The Wider Benefits of Learning Papers: No. 3. London: Institute of Education

Harmon, C. and I. Walker (2000) The returns to the quantity and quality of education: Evidence for men in England and Wales, *Economica*, 67, 19-35

Hartog, J. and H. Oosterbeek (1998) Health, wealth and happiness: Why pursue a higher education? *Economics of Education Review*, 17, 245-256

Hauck, K. and N. Rice (2003). *Targeting the worst-off. A longitudinal analysis of mental health*, Centre for Health Economics, University of York, mimeo

Haveman, R. and B. Wolfe (1984). Schooling and economic well-being: the role of non-market effects, *Journal of Human Resources*, 19, 377-407

Heckman J., H. Ichimura, P. Todd, (1997), Matching as an econometric evaluation estimator: Evidence from evaluating a job training program, *Review of Economic Studies*, 64,605-654

Heckman J., J. Stixrud and S. Urzua (2004), "The effects of cognitive and non-cognitive skills on labor and behavioural outcomes", Paper presented at the *Research Conference in Honor of Mark Berger*, University of Kentucky, November 2004

Henk H. , D. Katselnick, K. Kobak (1996) Medical costs attributed to depression among patients with a history of high medical expenses in a HMO, *Archive of General Psychiatry*, 53, 899-904

Kendler, K., Kessler, R., Neale,M., Heath, A. & Eaves, L., (1993). The prediction of major depression in women: toward and integrated etiologic model, *American Journal of Psychiatry*, 150, pp1139-1148.

Kenkel D. (1991). Health behaviour, health knowledge and schooling, *Journal of Political Economy*, 99, 287-305

Kenkel D., D. Lillard and A. Mathios (2004) "The role of high school completion and GED receipt in smoking and obesity", Paper presented at the *Research Conference in Honor of Mark Berger*, University of Kentucky, November 2004

Lleras-Muney A. (2001) "The relationship between education and adult mortality in the US", Working paper, Department of Economics, Columbia University.

Lewinsohn, P., Hoberman, H. & Rosenbaum, M. (1988). A prospective study of risk factors for unipolar depression, *Journal of Abnormal Psychiatry*. 97, pp251-264

Machin S. and S. McNally (2004). The literacy hour, *Centre for the Economics of Education*, London School of Economics, mimeo

Marmot, M.G.; Smith, G.D.; Stansfeld, S.; Patel, C.; North, F.; Head, J.; White, I.; Brunner, E. & Feeney, A. (1991) 'Health inequalities among British civil servants: the Whitehall II study', *The Lancet*, 337: 1387–1393.

Mazure, C., Bruce, M., Maciejewski, P. & Jacobs, S., (2000). "Adverse life events and cognitive-personality characteristics in the prediction of major depression and antidepressant response." American Journal of Psychiatry. 157, pp 896-903.

Meltzer, H.; Gill, B.; Petticrew, M. & Hinds, K. (1995) 'The prevalence of psychiatric morbidity among adults living in private households', Office of Population Censuses & Surveys: HMSO.

Mortensen, P., E. Agerbo, T. Erikson, P. Qin and N. Westergaard-Nielsen, 2000, 'Psychiatric illness and risk factors for suicide in Denmark', The Lancet, Vol 355, 9-12.

Patel, A. and Knapp, M. (1998) *Costs of mental illness in England*, Centre for the Economics of Mental Health, *Mental Health Research Review*: 5 May.

Propper C. , K. Jones, A. Bolster, S. Burgess, R. Johnston and R. Sarker (2004) "Local neighbourhood and mental health: evidence from the UK", University of Bristol, CMPO, WP 04/099

Rutter, M., (1967). "A Children's Behaviour Questionnaire for Completion by Teachers: Preliminary Findings." Journal of Child Psychology and Psychiatry, Vol **8**, pp1-11.

Rutter, M., Tizard, J. and Whitmore, K. (1970) *Education, Health and Behaviour*. London: Longmans

Silberg, J., Rutter, M., Neale, M. and Eaves, L (2001) "Genetic moderation of environmental risk for depression and anxiety in adolescent girls." British Journal of Psychiatry. 179, pp 116-121

Smith, J., 1999, Healthy bodies and thick wallets: The dual relation between health and economic status, *Journal of Economic Perspectives*, 13, 145-166

Zubenko, G., Maher B., Hughes H., Zubenko W., Stiffler J., Kaplan B., & Marazita M. (2003) "Genome-wide linkage survey for genetic loci that influence the development of depressive disorders in families with recurrent, early-onset, major depression. Am. J. Med. Genet (Neuropsychiatr. Genet.), in press, 2003.

WHO (1999) World health Report, 1999: Making a difference, WHO, Geneva

Annex 1: Malaise score

I often have back-ache	0.243
I feel tired most of the time	0.190
I often feel miserable or depressed	0.115
I often have bad head-aches	0.146
I often get worried about things	0.327
I usually have great difficulty sleeping	0.121
I usually wake unnecessarily early	0.167
I wear myself out worrying about my health	0.030
I often get into a violent rage	0.046
People often annoy and irritate me	0.218
At time I have twitching of face/shoulders	0.072
I often suddenly become scared for no good reason	0.049
I'm scared to be alone	0.032
I'm easily upset or irritated	0.163
I'm frightened of going out alone	0.048
I'm constantly keyed up and jittery	0.040
I suffer from indigestion	0.130
I suffer from an upset stomach	0.097
I have poor appetite	0.038
Every little things gets on my nerves	0.027
My heart often race like mad	0.057
I often have bad pains in my eyes	0.040
I have trouble with rheumatism or fibrosis	0.041
I had a nervous breakdown	0.026
Malaise score	2.43
Depression indicator	0.067