Cardiff Economics Working Papers

Working Paper No. E2022/4

Does higher education matter for health?

Sisi Ji and Zheyi Zhu

February 2022

ISSN 1749-6010

Cardiff Business School
Cardiff University
Colum Drive
Cardiff CF10 3EU
United Kingdom
t: +44 (0)29 2087 4000
f: +44 (0)29 2087 4419
business.cardiff.ac.uk

Does higher education matter for health?

Sisi Ji[*] Zheyi Zhu[†]

Abstract

Using 6 sweeps from 1958 British NCDS data we adopt a quasi-parametric approach of propensity score matching to estimate causal effects of higher education attainment on a wide range of cohorts' health-related outcomes at ages 33, 42 and 50. The non-pecuniary benefits to HE attainments on health are substantial. Higher educated cohorts are more likely to report better health, maintain a healthy weight, be non-smokers and to have a higher sense of control on drinking alcohol and are less likely to be obese. We also highlight the importance of investigating incremental returns to HE within the lifetime of cohorts. Effects on self-reported health (SRH), BMI, drinking alcohol increase with age but continuously decrease with smoking frequency. When taking into account gender heterogeneity, HE has a larger effect on BMI and likelihood of being obese for males and a greater effect on SRH and drinking alcohol and smoking frequencies for females. Furthermore, we find no significant evidence that HE reduces the likelihood of depression, both for males and females.

JEL Classification Codes: C21, I12, I23, I26

Keywords: Casual effect; Health; Higher Education; Propensity Score matching

[*] Cardiff Business School, Cardiff University, Cardiff, CF10 3EU, UK, Jiss@cardiff.ac.uk
[†] Cardiff Business School, Cardiff University, Cardiff, CF10 3EU, UK, Zhuz5@cardiff.ac.uk

1. Introduction

Education as a way of increasing human capital is a basic factor in the growth process of the aggregate economy. Although predominant studies confer most of the benefits that are likely to be reflected by the pecuniary return since the birth of the human capital theory (Schultz, 1961), bestows also give rise to a wide range of non-pecuniary benefits that could also consist in direct additions to welfare possibilities in terms of better health, longer life expectancy, less criminal behaviour, stronger social cohesion and greater political participation. In particular, educational attainment has been found to have a positive association with various health outcomes: the so-called "health education gradient" in decades of research (Grossman, 2006). According to Cutler, et al. (2006), education-health gradients increase when there is knowledge and technology available to prevent or treat because there is a universal demand for better health and those with more education, income, or power are likely to use new knowledge and new techniques more rapidly and effectively disease (Glied and Lleras-Muney, 2003; Cutler, et al., 2006).

The wider interests stem from the fact whether a true causal effect of education on health exists, then the individual's educational attainment represents the most obvious means through which policymakers could affect their health (Braga and Bratti, 2012). Individuals with high levels of education have invested to protect themselves by taking preventative measures to increase the probability of better health; hence, higher educated people tend to have better health than those with lower levels (Saxton, 2000). Although health education gradient may result in part from reciprocal causal effects between educational attainment and health status, other researches suggest that education does indeed have a causal effect on health (Currie and Moretti 2003; Wolfe and Zuvekas, 1997). The standard least square estimation may only represent simple correlations and face endogeneity problems, most scholars use the instrumental variable (IV) strategies or regression discontinuity (RD) designs to identify causal effects (Adams, 2002; Clark and Royer, 2013; Glied and Lleras-Muney, 2003; Jürges et al.2009; Meghir et al., 2018). These studies usually differ in terms of econometric specifications and focus only on single or very few health outcomes and behaviours at a particular age.

In this paper, we seek to add contributions to the existing literature in three main respects. First, the goal of this paper is to construct an estimate of the causal effect of higher education on health outcomes in the UK across the ages of 33, 42 and 50. The treated group includes the individuals who had completed some form of HE attainments and the control group includes the individual's highest education qualification is at least one A-level but not continue any

university studies. By including extensive covariates for family background characteristics, personal abilities and health status in childhood and adolescence, we characterise effects commonalities and compare the changes of the health returns of HE and the return to gender differences in the medium and long term by concentrating on a cohort who were continuously full-time employed during the period from 1991 to 2008, and therefore investigate whether the return gap between genders still exists when the cohorts are up to the age of menopause.

Second, we identify and estimate the treatment effect of HE on health and health-related behaviours using the propensity score matching (PSM) methodology (Rosenbaum and Rubin, 1983) which is widely applied in statistics and medical literature in both theoretical and empirical works (Heckman *et al.*, 1997; Dehejia and Wahba, 1999), in evaluating labour market policies (Lechner, 2002; Sianesi, 2004), assessing the effect of college quality (Berg Dale and Krueger, 2002, 2014; Black and Smith, 2004; de Luna and Lundin, 2014), and the wage return to education (Battistin and Sianesi, 2011; Blundell et al., 2003). The causal effect of the treatment is defined as the change in health outcomes caused by a potential move from untreated to treated status, or *vice versa*. Here, we focus on assessing the average treatment effects on treated assignment (ATT), which is the premium if individuals have been obtained HE attainment relative to their counterparts (non-HE attainment). It facilitates comprehensive evaluations of employing balance test to check the satisfaction of conditional independence assumption (CIA), a "thick-support" region test (Black and Smith, 2004) to check the estimates robustness, and associated Rosenbaum Bounds to check the satisfaction of selection on observable assumption.

Third, we use the National Child Development Survey (NCDS) data that can provide richer data sources on health and health-related variables. We therefore consider a wider set of health variables, in particular (i) general health outcome: self-assessed health; (ii) body weight health outcomes: Body Mass Index (BMI) and obesity; (iii) health-related damaging behaviours: frequency of smoking and drink alcohol; (iii) mental health outcome: depression based on malaise score. All of these health and health behaviours outcomes together provide a more general assessment of the effect of education on health.

The structure of the rest of this paper is as follows. Section 2 reviews related literature. Section 3 describes the method of PSM, empirical model and data description. The main empirical results are presented and discussed in Section 4. Section 6 highlights the main findings and draws the conclusion.

2. Literature review

The non-monetary benefits to education were posited in the very earliest works on human capital (Becker, 1964, Schultz, 1961). From an education perspective, the strength of this relationship suggests that health could be one of the most important sources of non-monetary returns to education. However, researches on whether the well-established striking relationship between education and health depends on a causal mechanism are not always clear-cut. Education could improve health through at least five channels[1] that have been proposed in the existing literature (Lochner ,2011). These channels provide evidence of the causal effects of an individual's education on a very wide set of health variables.

(a) The productive efficiency argument (Grossman, 1972) proposes that education directly affects the health production function. Given the same quantity of inputs, more educated individuals produce a higher stock of health than less educated ones. For example, education may impart direct knowledge about health and health behaviours, thereby shifting the health production function.

(b) The allocative efficiency argument (Grossman, 2006, Rosenzweig and Schultz, 1983) proposes that education has no impact on health unless it changes inputs in the health production function. The main mechanism through which education can affect the inputs is by increasing health-related knowledge.

(c) Changing time preference (Becker and Mulligan, 1997; Fuchs, 1982) so that individuals with a high discount rate are likely to be impatient, more likely to invest less in education, and more likely to engage in health-damaging behaviour. Hence, there could be a negative correlation between education and smoking which stems from an unobserved variable that does not reflect a true causal relationship.

(d) Changing economic factors resulting in higher levels of income and higher labour market opportunities allow individuals to work in less stressful jobs (Case and Deaton, 2005; Cutler and Lleras-Muney, 2010) and more highly educated people may tend to work in safer environments (Cutler and Lleras-Muney, 2008).

(e) Changing behavioural patterns including diet, smoking, obesity, patterns of alcohol consumption, preventative care, etc. (Mackenbach et al., 2008).

The available shreds of evidence on the causal effect of education on health are controversial

[1] Lochner (2011) lists the channels through which education may improve health with other identifications: stress reduction, better decision making or better information gathering, higher likelihood of having health insurance, healthier employment, better neighborhoods and peers and healthier behaviors.

in the UK (Jürges et al.,2009, 2011; Oreopoulos, 2006; Silles, 2009). Researches focus on an individual's general health status usually measured through self-reported heath (SRH) measures[2] or biomarker indicators[3]. Using compulsory schooling law changes as instruments, Oreopoulos (2006) applies an IV regression approach[4] based on the General Household Surveys (GHS) and identifies a positive and significant effect of education on SRH. Using age left full-time education as the measurement, the study finds a negative effect of education on physical and mental disability. Similarly, Silles (2009) using the same method based on data from Health Surveys of England and finds a positive causal effect of education (year of schooling) on SRH, which is much larger than the OLS estimates. The author further indicates that a strong health gradient is observed for other health measures, such as SRH and smoking behaviour. Using British Household Panel Survey, Contoyannis et al (2004) divide participants into 4 groups (degree, A-level, O-level, no qualification) by their maximum educational attainment. The authors apply Maximum Simulated Likelihood for a multivariate Probit model and find that educational attainment to self-rated health gradient remains significant, even after the inclusion of controls for lifestyles in the estimation and controlling for unobserved heterogeneity.

By contrast, Jürges et al. (2013) assess the causal link of compulsory schooling and health using two nationwide law changes in the minimum school leaving age in the UK as an exogenous variation for education. Their result shows there is no causal effect between compulsory schooling and the two biomarkers.[5] The impact of education on SRH is only significantly positive among the older female cohorts but was negative among younger female cohorts. The effect is insignificant among men across ages. Clark and Royer (2013) study the changes in the duration of compulsory schooling in the UK and find insignificant evidence of health returns in terms of improved health outcomes or changed health behaviours. The health outcomes they used are objective health measures, such as blood pressure, BMI, and levels of inflammatory blood markers.

Education to some extent induces individuals to have healthy lifestyles. Sabates and Feinstein (2004) propose a probit model based on data from the British Household Panel Survey to assess the relationship between education and health, particularly the uptake of health services in the

[2] It is argued self-reported measures may suffer from a variety of biases. An alternative unbiased measure is to use the objective biomarker indicator. This is because biomarker is a medical indicator allowing characterizing a biological processes as normal or pathological or requiring a pharmacologic intervention.

[3] However, in practice, such information is rarely available. Researchers usually use other health indicator as biomarker indicator, such as BMI, hypertension or chronic conditions.

[4] In particular, the author adopts the regression discontinuity method involving comparisons at the quarter-of-birth level. A regression discontinuity design can mitigate policy changes concerns by exploiting sharp changes in educational attainment.

[5] They are blood fibrinogen and blood C-reactive protein, respectively.

UK. The evidence finds that education has a direct effect on preventative health by raising awareness of the importance of undertaking periodic health tests. It favours a mechanism by which education increases the individual's self-efficacy and confidence, while also improving access to health services by increasing the individual's patience and motivation. The impact is still significant and robust after controlling factors such as income, social-economic status, and personal life circumstances. Cutler and Lleras-Muney (2010) report that by controlling for age, gender, and parental background, higher educated individuals in the US and UK[6] are less likely to smoke, less likely to be obese and less likely to be heavy drinkers; on the other hand, they are more likely to drive safely, more likely to live in a safe house, and more likely to use preventative care. In particular, for the UK, individuals with an A-level qualification are 12 % less likely to be smokers than less-educated individuals and 4 % less likely to become obese. This evidence is however in opposition to Clark and Royer (2013), who show no evidence that education improves behaviours in terms of the dietary regime and regular physical activity in the UK.

HE attainment is associated with greater income, more control over the working life, and with more varied and challenging work, and thus reduced morbidity (Marmot et al., 1991) but also higher levels of stress (Rose, 2001). Bynner et al. (2002) study a wide range of benefits of HE based on NCDS and BCS. They find that graduates are generally less depressed and present a higher sense of wellbeing than those with lower educational attainment. Feinstein (2002), using data from the NCDS and BCS and matching methods, shows that controlling for childhood abilities, health and family background factors, women from the 1958 cohort with lower secondary education have a 6% lower likelihood of depression than women with no qualifications, while these effects for men are weaker. In general, the results show that differences between individuals with different qualifications are substantially eroded when the selection bias is dealt with using matching methods. Chevalier and Feinstein (2006) rely on the NCDS dataset to control for childhood determinants and measures of mental health over the individual's life span to account for possible endogeneity of education. They estimate that individuals with at least O-levels reduce their risk of adult depression by 6 %. This effect is similar for men and women. However, Russell and Shaw (2009) focus on HE students in the UK and point out that a significant proportion of students studying in higher education present social anxiety, of which 10% of students are marked to have severe social anxiety. Nonetheless, this study does not identify a causal effect.

[6] In the UK case, they use data from Health & Retirement Study (HRS), Survey on Smoking (SOS), and NCDS to collect different health outcomes, and demographic and economic controls.

Studies on the effect of higher education on general health status disparities have rarely been found in the literature by adopting PSM or matching related approaches. Conti et al (2010) go beyond the existing literature which typically estimates mean effects to compute distributions of treatment effects and apply the matching method to show how the health returns to education can vary among individuals who are similar with respect to their observed characteristics. Based on a positive correlation between health and schooling conclusion, they then estimate causal effects of education (year of schooling) on adult health and healthy behaviours in a form of matching using the British Cohort Study in 1970. They conclude education has an important causal effect in explaining differences in health behaviours (such as smoking and regular exercise) as well as on some other outcomes (such as obesity poor health and depression). Besides that, they also show that family background characteristics, and cognitive, non-cognitive, and health endowments developed by early ages, are important determinants of the labour market and health disparities at age 30. Rosenbaum (2012) used data from the National Longitudinal Study of Adolescent Health to compare young adults ages 26 to measure the effect of highest degrees on measures of hypertension, obesity, smoking, sleep problems, and depression. The method they applied is the nearest-neighbour Mahalanobis matching within propensity score callipers. After matching, they found participants with baccalaureate degrees were 60% less likely to smoke daily, 14% less likely to be obese, and 38% less likely to have been diagnosed with depression.

3. Data and methodology

3.1 Causal inference

In econometrics evaluation studies, observational studies use a randomised trial to obtain an objective causal inference. However, data often does not come from randomised trials but non-randomised observations. Suppose an experimental design where the assignment to the case of a binary treatment is determined by a purely random mechanism:

$$D \perp X_{all} \tag{1}$$

where $D = \{0, 1\}$ is the indicator of exposure to treatment and X_{all} is the multidimensional vector of all observable and unobservable pre-treatment characteristics (covariates). In addition,

$$D \perp Y(0), Y(1) \tag{2}$$

The potential outcomes are then defined as $Y(D)$. This guarantees that D is independent with both observable and unobservable, and the potential outcomes will be statistically independent of the treatment status. With a randomised assignment, all of the characteristics of the

individuals are equally distributed between treated and untreated groups, which implies:

$$E(Y(0)|D = 1) = E(Y(0)|D = 0) \quad (3)$$

The causal effect for an individual unit *i* can be defined as the difference between the potential outcome in case of treatment and non-treatment:

$$T_i = Y_i(1) - Y_i(0) \quad (4)$$

where $i = 1,2 \ldots N$ and N denotes the total population. The evaluation problem arises because only one of the potential outcomes is observed for each individual i. The unobserved outcome is called a counterfactual outcome. Thus, the true causal effect of a treatment on individuals not subjected to the treatment can never be identified. The impossibility of observing both treatment and control outcomes for each individual is often referred to as the "fundamental problem of causal inference" (Rubin, 1978, Holland, 1986).

Hence, estimating the individual treatment effect T_i is not possible without making generally untestable assumptions and one has to concentrate on average treatment effects at the population. The average treatment effect (ATE) is defined as:

$$T_{ATE} = E[Y(1) - Y(0)] \quad (5)$$

Heckman (1997) argued that ATE might not be of relevance to policymakers because it includes the effect on persons for whom the programme never participated. One also concentrates on ATEs at a sub-population. The parameter of interest in most evaluation studies is then considered as the average treatment effect on treated (ATT), it is then defined as:

$$T_{ATT} = E(Y(1)|D = 1) - E(Y(0)|D = 1) \quad (6)$$

The problem is that $E(Y(0)|D = 1)$ is a hypothetical outcome because it is not observable and it depends on counterfactual outcomes. It then allows us to estimate the ATT by using:

$$[E(Y(1)|D = 1) - E(Y(0)|D = 0)] \quad (7)$$

In the absence of an experimental design or observational studies, using the mean outcome of untreated individuals $E(Y(0)|D$ is not recommended because it is likely that the covariates which determine the treatment decision also determine the outcome variable of interest. Thus, the differences in means between treated and untreated units would differ, even in the absence of treatment leading to a self-selection bias. For ATT this can be noted as:

$$E[Y(1)|D = 1] - E[Y(0)|D = 0] = T_{ATT} + E[Y(0)|D = 1] - E[Y(0)|D = 0] \quad (8)$$

where $E[Y(0)|D = 1] - E[Y(0)|D = 0]$ is defined as the self-selection bias.

3.2 The method of Propensity Scoring Matching

The idea of PSM methodology attempts, in a non-experimental context, to replicate the setup

of a randomised experiment. It is to match treatment and control units based on similar values on the propensity score $p(X)$, and the discarding of all unmatched units (Rubin, 2001). Ensuring PSM estimators identify and consistently leads to three assumptions:

1. Balancing of pre-treatment variables: $D \perp X \mid p(X)$
2. Unconfoundedness[7] or CIA: $D \perp Y1, Y0 \mid p(X)$
3. Common Support or overlap condition: $0 < p(D = 1|p(X)) < 1$

The PSM estimator for ATT takes the form:

$$T_{ATT}{}^{PSM} = E_{p(x)|D=1}\{E[Y(1)|D = 1, p(x) - E[Y(0)|D = 0, p(x)]\} \quad (9)$$

A general class of estimators of equation (9) can be written as:

$$T_{ATT}{}^{PSM} = \frac{1}{N^T}\sum_{i\in T} d_i Y_i - \frac{1}{N^C}\sum_{i\in T}(1 - d_i)w_j Y_i \quad (10)$$

Where N^T and N^C are the number of treated and untreated observations, w_j is a weight-related to the function of the estimated propensity score. The PSM estimator is the mean difference in outcomes over the common support, appropriately weighted by the propensity score distribution of participants (Caliendo and Kopeinig, 2008).

The advantage of the PSM approach is that it doesn't require the assumption of constant additive treatment effects across individuals. Heterogeneous treatment effects are permitted and can be retrieved via sub-group analysis, whereas standard parametric approaches assume homogeneous treatment effects across the sample analysed. This involves selecting the sub-group of interest and re-matching within that group and makes PSM a flexible tool for studying programme effects on groups of particular interest. PSM estimates of treatment effects are confined to counterfactuals in the area of common support and therefore do not rely on extrapolations beyond this region (Peel and Makepeace, 2012).

The PSM estimate compares the targeted outcomes of treated units with one or more control units. Matching algorithms are used to reduce bias by maximising the statistical similarities between treatment and non-treatment groups, while others maximise the number of matches to reduce variance by allowing comparisons between less similar treatments and control individuals (Rosenbaum and Rubin, 1985). In our model, we classify the treatment group (HE) as an individual's entry to all forms of higher education, including diploma, degree level, and higher degree level, and a control group (non-HE) who obtained one or more A-levels but who did not proceed into HE. Two matching methods: nearest neighbour matching with replacement and Kernel matching are considered to apply to the empirical model. The matching algorithms

[7] Rosenbaum and Rubin (1983) defined the assumption as Uuncofoundedness or ignorable treatment assignment.

are described in Appendix B.

3.3 Econometric model

The empirical model takes the following specification:

$$H = C + \beta HE + X\theta + \mu \quad (11)$$

where H is the measured outcome of an individual's general health, health behaviours and mental health. HE is the binary variable that stands for whether an individual obtains HE attainment. β is the parameter of interest, which measures the treatment effect of HE on the particular measure of health status and health-related behaviour. C is the constant term and μ is the error term. X is a vector of confounding variables before the HE decision that can explain variations both in treatment and outcomes variables but themselves are not inversely caused by treatments or outcomes.

3.4 Data

The British NCDS 1958 used in this paper is a continuing panel survey of all individuals born in the UK between the 3rd and 9th week of March 1958. There have been 10 follow up sweeps surveys available since birth up to 2020. In this study. we use the data from sweeps of all cohort members and their immediate families collected at ages 7, 11, 16, 33, 42 and 50.

One of the main advantages of using the NCDS is it allows us to account for the full information on the cohorts' contemporaneous characteristics, such as early cognitive ability, early parental information, educational attainment and subsequent working life. For educational attainment, it contains detailed information on the HE qualifications achieved by each individual up to 2000 and can be used to identify the type of qualification obtained and the information from the 1978 school exams file in the NCDS on school qualifications. We define HE attainment in the UK context as the return from undertaking some form of university level or equivalent. Here, we assume that individuals stop having further education in 1991 at the age of 33. The overall sample includes 1,444 individuals who have a HE qualification and 1,198 individuals who obtained at least one A-level but who did not continue into HE.

Table 1 Sample size of treat and control groups

	1 or more a-levels	HE degree	Total Sample
Men	523	782	1350
Women	675	662	1337
All	1198	1444	2687

3.4.1 Health outcomes

Six health-related outcomes are chosen across different ages, including indicators of general health status: SRH, measured BMI and obesity; indicators of healthy behaviours: alcohol drink frequency, smoking frequency and the indicator of mental health status. SRH is a subjective indicator of health that individuals assess relative to a representative person of the individual's own age. In NCDS, it measures how they feel about their health by using four categories: excellent, good, fair, and poor. I recode SRH so that a higher number corresponds to better SRH (i.e. 1 = poor, 4 = excellent). Figures 1 and 2 illustrate the distribution of SRH for different age levels by gender.

Figure 1 SRH for Men across different ages

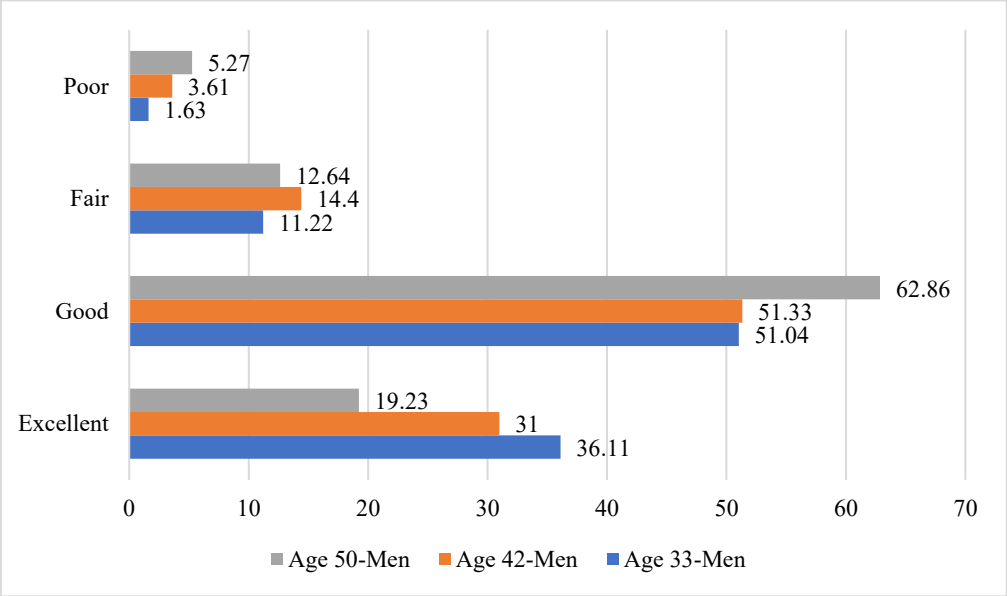
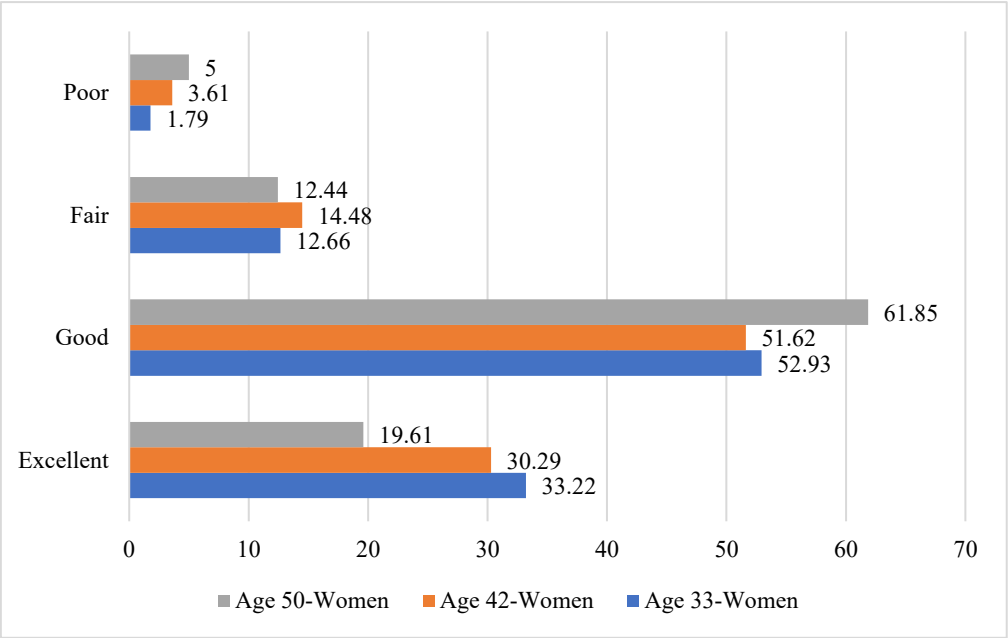


Figure 2 SRH for Women across different ages

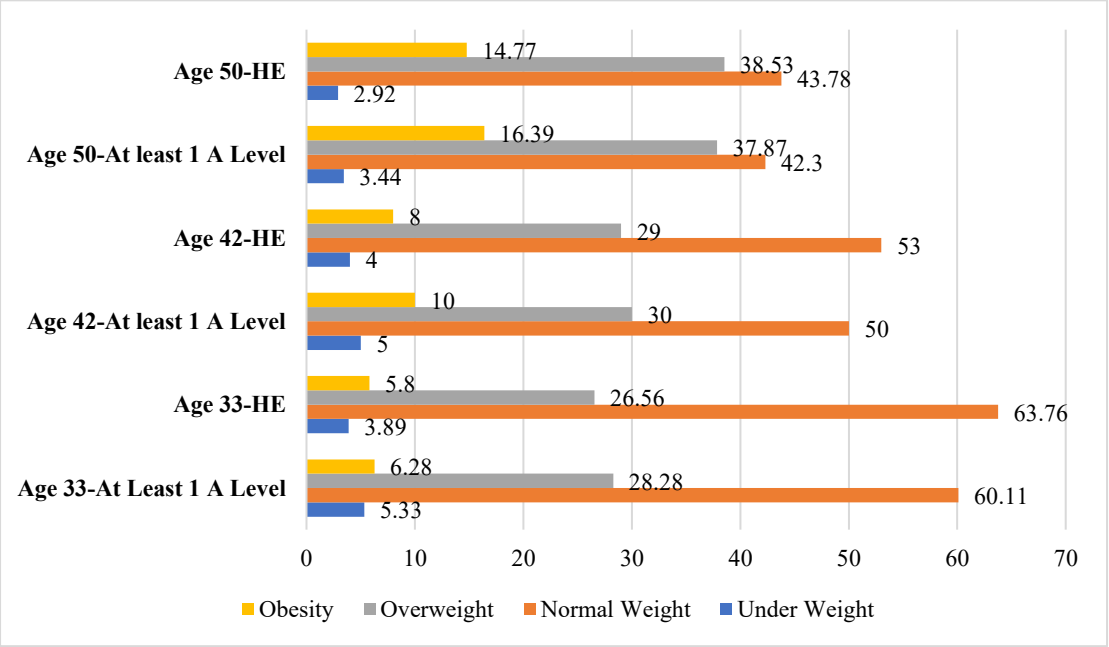


BMI is a useful measure of being overweight and obese, it is an estimate of body fat and is a gauge of the risk of diseases that are associated with more body fat. The NCDS records the height and weight of the respondents at all sweeps[8], except for Sweep 7 in 2004. Table 2 summarises the descriptive statistics of BMI. The measures of BMI can also be used to construct an indicator of being overweight or obese. According to the classification from World Health Organization (WHO), we place measured BMI into four categories, which are: underweight, normal weight, overweight, and obesity.[9]

Table 2 Descriptive statistics of BMI by gender over time

		Mean	S.D
Age 33	**Men**	25.39	4.01
	Women	23.68	4.38
Age 42	**Men**	26.28	4.21
	Women	25.27	5.08
Age 50	**Men**	27.52	4.63
	Women	25.50	4.77

Figure 3 Percentages of obesity by qualifications



[8] We use the following formula to calculate the respondents' BMI: $\text{BMI} = \frac{\text{weight(kg)}}{[\text{height(m)}]^2}$ or $\text{BMI} = \frac{\text{weight(lb)}}{[\text{height(inch)}]^2} \times 703$

[9] WHO classification can be found at: http://www.who.int/mediacentre/factsheets/fs311/en/

Figure 4 Percentage of drinking alcohol frequency across ages

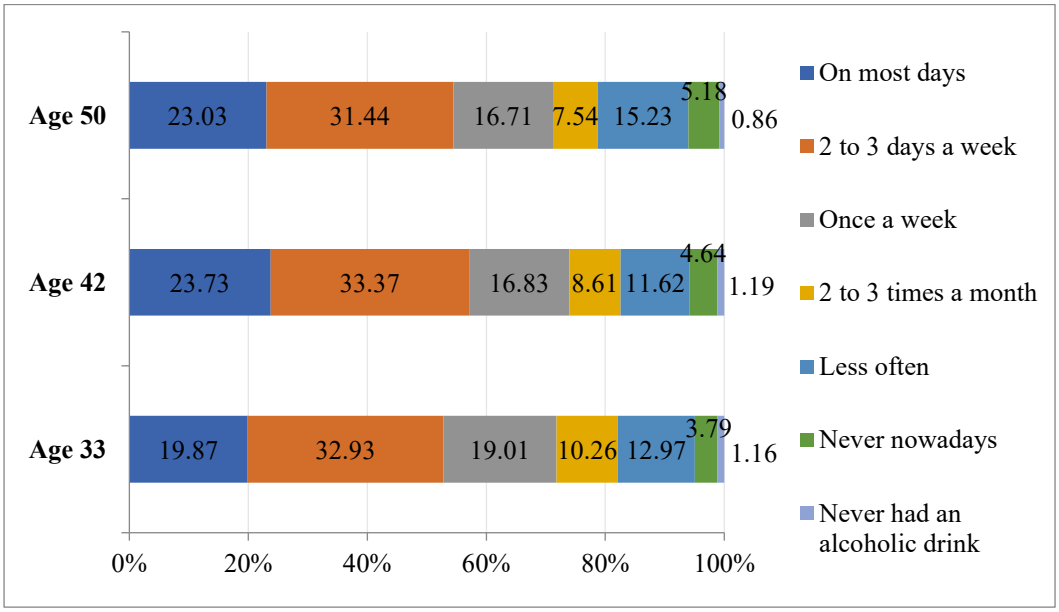
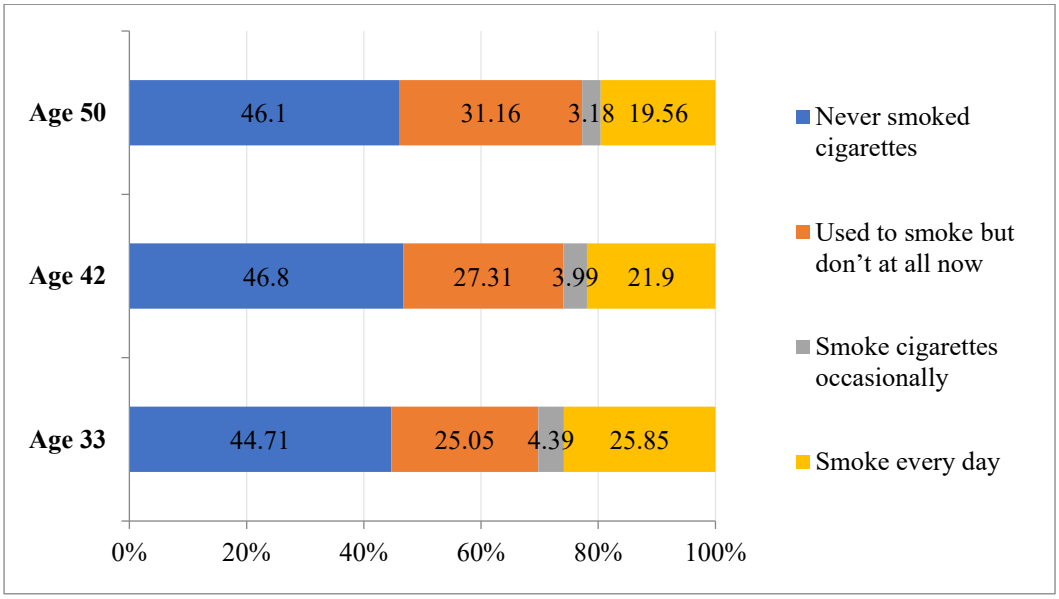


Figure 5 Percentage of smoking frequency across ages

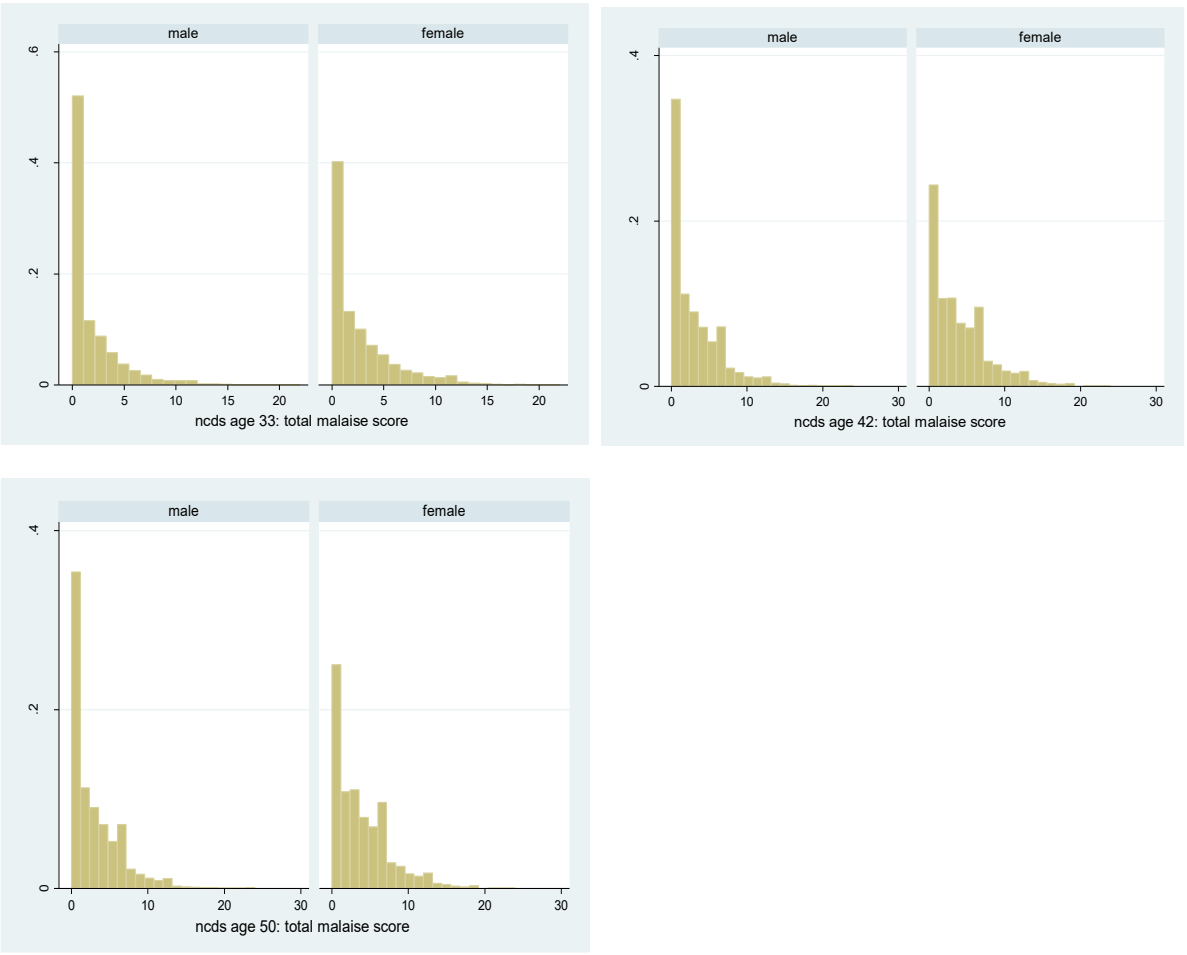


Figures 4 and 5 illustrate distributions of drinking and smoking frequency. The two data are directly collected from NCDS. The malaise score is calculated from the Malaise Inventory[10] and designed to identify depression in non-clinical settings and indicator of depression (Rutter et al., 1970). Figure 6 illustrates the distribution of malaise scores at different ages. According to the classification defined in NCDS, individuals responding 'yes' to eight or more of the 24

[10] It is a set of 24 self-completion questions combined to measure levels of psychological distress, or depression. The 24 'yes-no' items of the inventory cover emotional disturbance and associated physical symptoms, thus the score ranges from 0 to 24.

items are considered to be at risk of depression. We create a binary variable indicating depression = 1 when malaise score ≥ 8 and no depression = 0 when malaise score < 8.

Figure 6 Distribution of the malaise score at different ages



3.4.2 Confounding variables

Table 3 presents the descriptive statistics for confounding variables. Confounding variables are considered to influence both the educational decision and health outcomes should be included as regressors. The choice of cofounding variables is dictated into five main categories.

1. Demographic characteristics: region of residence at birth, ethnicity.
2. School type: type of secondary school.
3. Early age personal ability: Mathematics score, reading score accessed at age 7 and 11.
4. Family backgrounds: Father's year of education and father's social class, mother's year of education, mother's employment status number of siblings, parents' interest in participant's education, all at age 16, and family finance status at age 11 and 16.
5. Health status in childhood and adolescence.

Table 3 descriptive statistics for counding variables

Variable	Mean	S.D	Variable	Mean	S.D
White	0.987	0.115	**Father's social class in 1974**		
Mathematics ability at 7 years			Professional	0.034	0.181
5th quintile (highest)	0.194	0.395	Intermediate	0.132	0.338
4th quintile	0.114	0.318	Skilled Non-manual	0.063	0.244
3rd quintile	0.271	0.445	Skilled manual	0.298	0.458
2nd quintile	0.141	0.348	Semi-skilled non-manual	0.01	0.01
1st quintile (lowest)	0.28	0.449	Semi-skilled manual	0.087	0.281
Reading ability at 7 years			Unskilled	0.036	0.185
5th quintile (highest)	0.192	0.394	Missing, or unemployed or no father	0.34	0.474
4th quintile	0.132	0.339	Number of siblings in 1974	1.743	1.512
3rd quintile	0.263	0.44	**Father's interest in education**		
2nd quintile	0.209	0.407	Expects too much	0.024	0.153
1st quintile (lowest)	0.204	0.403	Very interested	0.262	0.44
Mathematics ability at 11 years			Some interest	0.249	0.433
5th quintile (highest)	0.194	0.396	**Mother's interest in education**		
4th quintile	0.202	0.402	Expects too much	0.037	0.188
3rd quintile	0.171	0.376	Very interested	0.349	0.477
2nd quintile	0.202	0.401	Some interest	0.346	0.476
1st quintile (lowest)	0.231	0.422			
Reading ability at 11 years			**Region in 1974**		
5th quintile (highest)	0.159	0.365	North West	0.116	-0.32
4th quintile	0.191	0.393	North	0.075	0.264
3rd quintile	0.241	0.428	East and West Riding	0.087	0.281
2nd quintile	0.168	0.374	North Midlands	0.076	0.265
1st quintile (lowest)	0.241	0.428	East	0.086	0.28
Comprehensive school 1974	0.467	0.499	London and South East	0.16	0.367
Secondary modern school 1974	0.17	0.376	South	0.063	0.243
Grammar school 1974	0.087	0.281	South West	0.068	0.251
Private school 1974	0.06	0.214	Midlands	0.101	0.301
Other school 1974	0.017	0.13	Wales	0.058	0.234
Father's age in 1974	46.641	6.39	Scotland	0.111	0.315
Mother's age in 1974	43.561	5.7	Other	0.1	0.299
Mother employed in 1974	0.657	0.475	**Father's years of education**	7.904	1.622
Bad finances in 1969 or 1974	0.114	0.317	**Mother's years of education**	7.916	1.376

4 Empirical analysis

The propensity score is estimated by probit regression of the binary treatment variable on a vector of confounding variables identical to that in Table 3. Results are reported in table A.1 in Appendix A. Distributions of propensity scores and common support regions for the treated and control groups are displayed in figure A.1 to ensure the validity of the overlap assumption. Using the different ages as reference points, the results presented are disaggregated by various educational groups. It is noted the full sample size for both genders is not always equal to the sum of the male and female sub-samples because pooling the samples leads to different matches to those in the sub-samples. OLS results are reported together with PSM estimates based on two different matching algorithms discussed. Although suffering endogeneity bias, the parameter of interest in OLS estimates empirically can be interpreted as the average treatment effect (Aizer et al., 2016, Voigtländer and Voth, 2012) and be used to find how the health behaviours change overtime with full controls of confounding variables. We also consider it may not be appropriate to show the percentage change of the treatment effects of He on ordered categorical outcomes. After matching, we tabulate the total matched sample and calculate the fraction of each ordered categorical outcome for both treatment and control groups. It is a straightforward measure of percentage change of the treatment effects across ages.

4.1 Self-reported health

As shown in table 4, OLS estimates for the age group of 33 is about 0.064 for the whole sample, 0.037 for male, and 0.079 for female participants. In general, the effect for age 42 and 50 are somewhat similar to that of age 33, HE has a greater effect for females rather than males. All of the ATTs estimated are statistically significant at the 90% level. The effects on the pooled sample across ages have no significant differences: 0.08 at age 33, 0.08 at age 42, and 0.09 at age 50. The results stress the importance of taking sex heterogeneity into account while the effects show monotonic increases with age. Sub-sample analysis by gender further indicates that this result is significant for females where the effect size has a 0.03 margin more than that of the male group at all ages. For males, individuals with HE attainment at age 33 enjoy an extra 0.079 margin on SRH, 0.09 at age 42, and 0.1 at age 50, respectively. Females enjoy an extra 0.04 margin at age 33, 0.045 at age 42, and 0.067 at age 50.

Table 4 Causal effects of HE on self-reported health

Age	Baseline OLS			PSM NN			PSM Kernel		
	Full	Male	Female	Full	Male	Female	Full	Male	Female
33	-0.0639**	-0.0365	-0.0786	-0.081*	-0.078*	-0.118**	-0.078*	-0.0704*	-0.111**
	(0.142)	(0.105)	(0.134)	(0.049)	(0.071)	-(0.075)	(0.034)	(0.054)	(0.067)
42	-0.0653	-0.0521	-0.0832	-0.085*	-0.090**	-0.135**	-0.081*	-0.090**	-0.131**
	(0.034)	(0.053)	(0.045)	(0.051)	(0.057)	(0.074)	(0.045)	(0.05)	(0.071)
50	-0.0671	-0.059	-0. 0856	-0.091*	-0.102*	-0.167**	-0.090*	-0.100*	-0.165**
	(0.054)	(0.083)	(0.072)	(0.07)	(0.075)	(0.076)	(0.072)	(0.064)	(0.069)

Note: **significant at the 5% level; *significant at 10% level

Table 5 shows 41.9% of males with HE attainments of the total treated sample size are categorised as excellent, whereas that of non-HE males are computed as 32.5% of the total untreated sample. This implies the impact of a HE is to increase the incidence of good health by 30 % points. On the other hand, when measuring the risk of poor health status, the risk is more than doubled from 0.9% (with HE) to 2.6% (with non-HE). For females, the friction of the 'excellent' category is relatively close (38.4% and 37.0%), whereas the risk of having poor health status also doubles from 1.3% to 2.8% if females do not obtain a HE attainment. The rest of the results also show substantive evidence to suggest that HE has a significantly positive impact on an individual's general health status in terms of SRH condition across the age. Higher educated cohorts have better general health conditions and this impact increases as cohorts get older. The results are somewhat consistent with the previous finding by Ross and Wu (1995), and White et al. (1999), which suggest that education has a strong and positive effect on adult SRH.

Table 5 Fraction of total matched sample, self-reported heath

	Male		Female	
	HE	**Non HE**	**HE**	**Non HE**
		Age 33		
Excellent	257 (41.9%)	77 (32.5%)	229 (38.4%)	107 (37.0%)
Good	321 (52.4%)	122 (51.4%)	323 (54.2%)	147 (50.9%)
Fair	29 (4.8%)	32 (13.5%)	36 (6.1%)	27 (9.3%)
Poor	6 (0.9%)	6 (2.6%)	8 (1.3%)	8 (2.8%)
Matched sample	613	237	596	289

		Age 42		
Excellent	228 (44.3%)	63 (32.1%)	182 (35.8%)	78 (32.9%)
Good	241 (46.8%)	99 (50.5%)	241 (47.4%)	110 (46.4%)
Fair	37 (7.2%)	25 (12.7%)	63 (12.4%)	29 (12.2%)
Poor	8 (1.7%)	9 (4.7%)	22 (4.4%)	10 (8.5%)
Matched sample	515	196	508	237
		Age 50		
Excellent	135 (26.9%)	37 (18.4%)	97 (20.3%)	40 (17.3%)
Good	205 (40.7%)	83 (41.3%)	208 (43.5%)	91 (38.8%)
Fair	125 (24.8%)	60 (30.0%)	132 (27.6%)	64 (27.3%)
Poor	39 (7.6%)	21 (10.3%)	41 (8.6%)	39 (16.6%)
Matched sample	504	201	478	234

3.5.2 BMI and obesity

The OLS estimates in table 6 show HE has a strong and negative association with BMI, which suggests a massive decrease in BMI when obtaining HE attainment. Estimates are significant at age 42 and 50, but insignificant at age 33. The size of the association for males is larger than females in absolute values. However, it is striking that in the pooled sample, the overall impact first increases at age 42 but decreases at age 50.

When turning to PSM estimates with the inclusion of full controls for covariates, the estimated coefficient from PSM has no significant difference compared to the OLS result at age 33 in pooled samples. HE appears to have a larger effect on reducing the BMI figure for males (0.356) than females (0.136) at age 33. However, except for males, none of these estimated coefficients is statistically significant. The HE reduces BMI figure up to 0.472 at age 42 and 0.617 at age 51 in pooled samples. As the cohorts grow older, males get more benefits from being highly educated to control the BMI figures. The figures are reduced by 0.529 at age 42 and 0.856 at age 50, respectively, almost twice as large as that of females.

We also consider the effects of HE on the threshold of obesity. The ATTs are insignificant when the cohorts are aged 33. Once cohorts grow to age 42, the marginal effects become -0.123 for males and -0.107 for females, both significant at the 95 % confidence level. The magnitude of the effect continues to slightly increase when individuals are aged 50, which accounts for -0.136 (males) and -0.114 (females). This implies that HE attainment have a significant but small restraining effect on obesity growth for individuals at age 42 and 50.

Table 6 Causal effects of HE on BMI and Obesity

	OLS			PSM NN			PSM Kernel		
	BMI								
Age	Full	Male	Female	Full	Male	Female	Full	Male	Female
33	-0.259	-0.342	-0.102	-0.297	-0.355*	-0.136	-0.301	-0.360*	-0.138
	(0.165)	(0.229)	(0.108)	(0.192)	(0.152)	(0.362)	(0.114)	(0.140)	(0.245)
42	-0.5462**	-0.5497**	-0.482**	-0.472**	-0.529**	0.377**	-0.475**	-0.528**	0.376**
	(0.181)	(0.24)	(0.147)	(0. 031)	(0.040)	(0.035)	(0. 115)	(0.124)	(0.103)
50	-0.3304	-0.4671*	-0.632**	-0.617**	-0.859**	-0. 481*	-0.601**	-0.821**	-0. 424*
	(0.206)	(0.279)	(0.273)	(0.242)	(0.364)	(0.127)	(0.211)	(0.301)	(0.114)
	Obesity								
33	-0.032	-0.071*	-0.041	-0.026	-0.064	-0.015	-0.024	-0.06	-0.015
	(0.028)	(0.04)	(0.044)	(0.032)	(0.076)	(0.06)	(0.029)	(0.070)	(0.061)
42	-0.075**	-0.108**	-0.087**	-0.110**	-0.123**	-0.107**	-0.101**	-0.119**	-0.100**
	(0.029)	(0.039)	(0.433)	(0.052)	(0.051)	(0.046)	(0.050)	(0.049)	(0.042)
50	-0.065*	-0.116**	-0.079*	-0.124**	-0.136**	-0.114**	-0.118**	-0.130**	-0.109**
	(0.035)	(0.047)	(0.049)	(0.064)	(0.059)	(0.045)	(0.061)	(0.048)	(0.039)

Note: **significant at the 5% level; *significant at 10% level

4.2 Drinking and smoking frequency

OLS results in table 7 show cohorts with HE attainment leads to a gradual reduction in alcohol consumption frequency from 33 to 50. The negative treatment effect indicates higher educated cohorts are likely to drink less alcohol. Such effect is still significant when taking into account gender heterogeneity. Both male and female respondents with HE are likely to drink less as they grow older. The estimated ATTs from PSM for males at age 33 is insignificant but when the cohorts grow older, the effect for men dramatically goes up to 0.156 at age 42 and 0.201 at 50, both statistically significant at 90% level. Females with HE shows a 0.255 margin compared to those without. The margin also has a remarkable increase to 0.416 at age 42 and 0.474 at age 50.

Likewise, the results for OLS show a positive impact of HE on the incidence of smoking. Corhots with HE attainment reduce ranging from 0.07 to 0.15 on average and the effect on smoking steadily decreases in the long term for both genders. The results for PSM estimates are mixed. The parameter of interest that shows the impact of HE on smoking at age 33 is reported about 0.15 for the pooled sample. Meanwhile, higher educated females are nearer to "never smoke" compared to males. Attending HE can significantly gain a 0.204 margin for females. By contrast, the effects are observed to be insignificant for males. When participants grow older, the impact goes down by 0.05 at age 42 for the pooled sample. On the female sub-sample, the marginal effect only accounts for 0.106, or almost half the figure compared to that

when they were 9 years younger. This effect for males is still insignificant. Furthermore, we do not find any significant effects of HE on reducing the frequency of smoking behaviour when the participants enter their 50s for both genders.

Turning to the fraction changes of each category for matched samples in table 8, males with HE are more likely to quit smoking than the ones without HE at age 33. Occasional smoking frequency for HE participants is less than that for Non-HE participants, whereas daily smoking frequency for both groups is almost the same. For females, the daily smoking frequency for the HE group is higher than that for the non-HE group, but the occasional smoking frequency does not have significant differences. Moreover, the quit-smoking friction of the non-HE group is higher than the HE group is because people in the HE group are more likely to be a non smoker. As the participants get older, the differences between the two groups become less. It is found that at age 50, the friction of four categories for both treated and control groups are almost equivalent.

Overall, these findings reinforce the findings by a number of previous studies which have found a negative correlation between smoking and education (Feinstein et al., 2008), and between drinking alcohol and education in the case of the UK (Cutler and Lleras-Muney, 2010). More educated young adults tend to hold risk perceptions more closely related to the actual risks of these behaviours. However, our result also suggested that the impact of HE is decreasing as individuals are getting older. In particular, HE does not effectively affect smoking behaviour when cohorts are in their age 50.

Table 7 Causal effects of HE on drinking and smoking frequency

					PSM NN			PSM Kernel	
		OLS							
Age	**Full**	**Male**	**Female**	**Full**	**Male**	**Female**	**Full**	**Male**	**Female**
	Alcohol Drinking Frequency								
33	-0.1779**	-0.0317	-0.2864**	-0.231**	-0.073	-0.255**	-0.214**	0.067	-0.245**
	-0.069	-0.096	-0.099	-0.034	-0.053	-0.045	-0.031	-0.05	-0.038
42	-0.2321**	-0.1384	-0.2615**	-0.301**	-0.156*	-0.416	-0.294**	-0.148*	-0.409*
	-0.076	-0.104	-0.111	-0.041	-0.024	-0.044	-0.034	-0.014	-0.039
50	0.2631***	0.1298	0.3615***	-0.358**	-0.201*	-0.474**	-0.345**	-0.194*	-0.456**
	-0.034	-0.106	-0.113	-0.034	-0.025	(0. 048)	-0.028	-0.025	-0.04
	Smoking Frequency								
33	-0.141**	-0.134**	-0.150**	-0.145**	-0.082	-0.204**	-0.141**	-0.08	-0.200**
	-0.041	-0.05	-0.057	-0.072	-0.07	-0.101	-0.072	-0.07	-0.101
42	-0.101**	-0.093*	-0.129**	-0.093**	-0.053	-0.106**	-0.088**	-0.048	-0.101**
	-0.042	-0.058	-0.058	-0.048	-0.109	-0.051	-0.048	-0.109	-0.051

50	-0.098**	-0.073*	-0.116**	-0.074	-0.046	-0.097	-0.071	-0.039	-0.089
	-0.039	-0.041	-0.053	-0.034	-0.057	-0.064	-0.034	-0.057	-0.064

Note: **significant at the 5% level; *significant at 10% level

Table 8: Fraction of total matched sample: drinking and smoking frequency

	Male		Female	
	HE	Non-HE	HE	Non-HE
	Drinking Frequency			
	Age 33			
Once a day	143 (26.5%)	105 (30.5%)	65 (14.6%)	85 (19.8%)
2 to 3 days a week	253 (46.9%)	173 (50.3%)	207 (46.5%)	199 (41.7%)
Once a week	57 (10.6%)	28 (8.1%)	72 (16.2%)	65 (15.2%)
2 to 3 times a month	46 (8.5%)	16 (4.7%)	51 (11.5%)	17 (4.0%)
Less often or only on special occasions	30 (5.6%)	16 (4.7%)	30 (6.7%)	45 (10.5%)
Never nowadays	7 (1.3%)	3 (0.9%)	15 (3.3%)	9 (2.9%)
Never had an alcoholic drink	3 (0.5%)	3 (0.9%)	5 (1.1%)	9 (2.9%)
Matched Sample	539	344	445	429
	Age 42			
Once a day	141 (29.9%)	104 (36.1%)	104 (24.6%)	110 (28.3%)
2 to 3 days a week	203 (43.0%)	126 (43.8%)	162 (38.3%)	147 (37.8%)
Once a week	67 (14.2%)	31 (10.8%)	70 (16.5%)	60 (15.4%)
2 to 3 times a month	22 (4.7)	19 (6.6%)	27 (6.4%)	25 (6.4%)
Less often or only on special occasions	34 (7.2%)	6 (2.1%)	37 (8.7%)	25 (6.4%)
Never nowadays	3 (0.6%)	1 (0.3%)	18 (4.3%)	17 (4.4%)
Never had an alcoholic drink	2 (0.4%)	1 (0.3%)	5 (1.2%)	5 (1.3%)
Matched Sample	472	288	423	389
	Age 50			
Once a day	181 (39.2%)	133 (43.8%)	106 (25.2%)	99 (26.4%)
2 to 3 days a week	147 (31.8%)	91 (30.0%)	145 (34.4%)	137 (32.4%)

Once a week	62 (13.4%)	34 (11.1%)	64 (15.2%)	53 (12.6%)
2 to 3 times a month	17 (3.7%)	13 (4.3%)	52 (12.4%)	25 (6.0%)
Less often or only on special occasions	52 (11.3%)	28 (9.2%)	40 (9.5%)	45 (10.7%)
Never nowadays	2 (0.4%)	1 (0.3%)	13 (3.1%)	12 (2.8%)
Never had an alcoholic drink	1 (0.2%)	0 (0.0%)	1 (0.2%)	4 (0.9%)
Matched Sample	462	300	421	375
	Smoking Frequency			
	Age 33			
Never smoke	359 (60.0%)	139 (59.6%)	355 (60.2%)	155 (54.8%)
Used to smoke	112 (18.7%)	35 (15.0%)	124 (21.0%)	68 (24.0%)
Smoke occasionally	36 (6.0%)	19 (8.2%)	20 (3.4%)	12 (4.3%)
Smoke everyday	97 (16.2%)	40 (17.2%)	91 (15.4%)	48 (17.0%)
Matched Sample	598	233	590	283
	Age 42			
Never smoke	310 (60.0%)	116 (58.6%)	310 (59.8%)	128 (54.2%)
Used to smoke	104 (20.1%)	40 (20.2%)	110 (21.2%)	60 (25.4%)
Smoke occasionally	42 (8.1%)	17 (8.6%)	39 (7.5%)	21 (8.9%)
Smoke everyday	61 (11.8%)	25 (12.6%)	59 (11.4%)	28 (11.9%)
Matched Sample	517	198	518	236
	Age 50			
Never smoke	313 (61.2%)	110 (59.3%)	296 (59.3%)	130 (54.6%)
Used to smoke	118 (23.1%)	48 (23.5%)	151 (29.3%)	64 (33.6%)
Smoke occasionally	35 (6.9%)	16 (7.8%)	13 (2.6%)	6 (2.5%)
Smoke everyday	45 (8.8%)	19 (9.3%)	44 (8.8%)	22 (9.2%)
	511	204	499	238

4.3 Depression

The OLS results find a negative relationship between HE and depression shown in table 9.

These associations vary significantly for different ages. HE has a larger impact on depression for females at age 33 than for males. For the PSM estimates, all of the estimated coefficients appear to be negative but insignificant, ATT is only significant for females at age 33. The PSM results suggest that most of the depression-education gradient in OLS comes from selection rather than causation. A general increase in the malaise score and depression indicator over time for both genders, but it appears that participation in HE does not carry potential causal effects.

In contrast to previous research evidence (Bynner et al., 2003; Feinstein et al., 2002), our finding does not suggest a significant impact of HE on the reduction in depression. This possibly arises because those studies mainly focus on those participants with lower or no qualifications. Lower educated individuals can benefit from education, and they may acquire better labour market opportunities or higher wages in return. As a result, they are more likely to work and have a better lifestyle and will be less likely to suffer from depression. However, the causal effect of HE on depression is ambiguous since there may be contrasting mechanisms. HE attainment is be associated with more control over working standards and thus has a positive effect on mental health and reduces rates of morbidity; on the other hand, higher occupational attainment also leads to higher levels of stress. It is believed that there could be important trade-offs between stress and satisfaction that may lead to a complex and non-linear relationship between educational success and mental health (Hartog and Oosterbeek, 1998).

Table 9 Causal effects of HE on depression

	OLS			**PSM NN**			**PSM Kernel**		
	Full	**Male**	**Female**	**Full**	**Male**	**Female**	**Full**	**Male**	**Female**
Age 33	0.097**	0.080**	0.113**	-0.007	-0.001	-0.026*	-0.006	-0.001	-0.021*
	(0.034)	(0.052)	(0.044)	(0.009)	(0.01)	(0.013)	(0.009)	(0.01)	-0.013
Age 42	0.082**	0.078**	0.120**	-0.011	-0.006	-0.073	-0.01	-0.006	(0.07)
	(0.05)	(0.045)	(0.097)	(0.049)	(0.026)	(0.093)	(0.049)	(0.026)	(0.093)
Age 50	0.104**	0.094**	0.123**	-0.018	-0.012	-0.107	-0.015	-0.01	-0.102
	(0.084)	(0.072)	(0.094)	(0.064)	(0.053)	(0.105)	(0.064)	(0.053)	(0.105)

Note: **significant at the 5% level; *significant at 10% level

4.4 Robustness test

Making causal claims about effects are considered to satisfy the three assumptions aforementioned. To further test the credibility of the estimated results, we conduct a thick region test and balance test and examine the sensitivity of the results due to unobserved

heterogeneity by Rosenbaum Bounds (R-bounds). To test the robustness of the PSM estimates, we follow Black and Smith (2004) and estimate the ATTs on the region of thick-support, which is defined as the region with an estimated propensity score in the interval by $0.33 < \hat{P}(X) < 0.67$. The authors adapted this approach based on two reasons. First, the fact that individuals with high estimated PSs observed at low levels of treatments may represent a measurement error in the treatment variable. Second, there may be a residual selection on unobservables which will have a large effect on the bias for values of the propensity score in the tails of the distribution. In practice, the 'thick-support' region is characterised by having a substantial number of observations in both the treatment group and the comparison group, which means that the average frequency with which a comparison observation is used as a match is comparatively low.

The estimated effects for the thick-support region thus refer to samples that, in terms of sheer size, are very different to those on the entire common support. In our case, imposing the thick-support condition leads to a drop of roughly one-third of the observations in the pooled samples. As presented in table A.2, the thick-support estimates in the majority of the cases seem fairly robust compared to the estimates based on the entire common support region. Although the estimates generally indicate a slight increase in the HE impact of health and health-related indicators, the estimated effects on the thick-support are similar to those on the entire common support, which is an indication of effect homogeneity over different values of the propensity score.

4.5 Balance test for matching quality

The adequacy of the matching process was evaluated by assessing covariate balance using mean absolute bias and Pseudo-R^2, as advocated by Caliendo and Kopeinig (2008). Table A.3 presents the covariate balance statistics concerning the joint quality of the matching before and after matching. The overall mean absolute bias before matching lies between 10 to 30 %. The matching generates a reduction in mean bias by approximately six times. After matching, the bias is significantly reduced for the NN and kernel matching estimators, ranging from 2 to 8 %. In particular, Kernel matching provides a better result and shows that all after matching covariates display a mean absolute lower than that from NN matching. On the other hand, Pseudo-R^2 indicates how well the covariates explain the probability of receiving treatment. The reported Pseudo-R^2 before matching is normally around 15 to 30 % whereas after matching it drops to roughly about 3 %. This indicates that there are fewer systematic differences in the

distribution of covariates between the treatment and the control groups. These results clearly show that the matching procedure is fairly successful in terms of balancing the distribution of covariates between the two groups.

4.6 Sensitivity analysis for unobserved heterogeneity

We apply sensitivity analysis based on the R-bounds to test the robustness of the results, as displayed in table A.4. The increasing bound parameter Γ would result in a statistically insignificant treatment effect if there is an unobserved heterogeneity. As discussed in Chapter One, starting from Γ= 1, i.e. there is no hidden bias.

We first assessed the effects on general health indicators in panel A. The sensitivity analysis shows that for effect on SRH, through the increase of Γ up to 1.10, the upper bound of the p-value exceeds the 5%-level, and this occurs to all matching algorithms. This indicates that the result is relatively vulnerable to unobserved bias, while it only requires a 10% increase in the odds of selection to negate the effect. Similarly, for effects on BMI and obesity, it would also generally take relatively low Γ values of unobserved selection (about 1.15 on average for the former and about 1.10 on average for the latter) to make the treatment effect statistically insignificant. Moreover, when I considered adding economic factors into covariates (log wage), the results were not improved significantly. For example, the P-value exceeds the 5%-level when Γ≈1.15 in the case of SRH. Therefore, the robustness of the effect on general health indicators needs further investigation.

When turning to health behaviour outcomes in panel B, the treatment effect would have been insignificant when Γ≈1.31 to 1.45. It is at this value that the treatment effect is no longer statistically significant at 5 %. This suggests that having matched on observed covariates, any unobserved confounding variables would have to increase the likelihood of selection by around 35 %. This is considered a fairly large value. In addition, in common with health behaviour outcomes, any unobserved confounding variable would have to increase the odd ratio by over 50% (Γ≈1.50) to overturn the causal effects on depression, as shown in panel C. Therefore, apart from the general health indicator, the estimated causal effects appear robust to the unobserved heterogeneity.

5 Conclusion

One weakness of the most existing evidence to date is that much of the assessment of the effects of education have measured education in terms of years of schooling. This has commonly been investigated as a simple linear effect, without distinguishing the relative benefit of educational participation at some particular stage. By using the longitudinal survey of NCDS data with different sweeps, this paper adopts a quasi-parametric approach of PSM to estimate the causal effects of HE attainment on a very wide range of cohorts' health-related outcomes across different ages. Individual's childhood cognitive ability, regions, secondary school types, parental information, health status in childhood, and adolescence have been taken into account as control variables to reduce the heterogeneity bias and measurement errors. Moreover, another key contribution is that we have also highlighted the importance of investigating whether there are incremental returns to HE within the lifetime of cohorts.

We draw the following conclusions from our empirical evidence. HE has a significantly positive impact on an individual's general health status in terms of self-assessed health status. Higher educated individuals have better general health conditions and this impact increases as the cohorts grow older. Evidence confirms a positive effect of education on obesity while higher education tends to have a lower BMI index. Such causal effects are significant when individuals are in their 40s and 50s. The HE also has substantial causal effects on initiation, cessation, and frequency of smoking and drinking alcohol, however, the effects on reducing the frequency of smoking are decreasing as cohorts are getting older.

In general, this paper suggests that attending HE is an effective way to improve general health status and reduce the likelihood of health-damaging behaviours. This finding is consistent with the fundamental causes of disease hypothesis (Link and Phelan 1995), which suggests that education gives an individual a wide range of serviceable resources, including money, knowledge, prestige, power and beneficial social conditions, which can be used to one's health advantage. Thus, a higher effect on an individual's health outcomes and health-related behaviours over time may be caused by the benefits of new effective techniques and the individual's confidence in curing disease, which has been built by having more knowledge. we support the view that education has a positive effect on an individual's health outcomes and reduces damaging health behaviours.

On the other hand, it is striking that the causal effect of HE on reducing the likelihood of depression in the UK is insignificant. This may happen because HE attainment results in a higher occupation in the labour market and this lead to higher levels of stress. There could be

existing trade-offs between stress and satisfaction of higher occupation that may lead to an ambiguous relationship between educational success and mental health.

we exhibit some robust evaluations and evidence of the quantitative effects of education assessed in terms of covariates balance and sensitivity to the unobserved hidden bias. The problem arises from the general health indicator. It indicates that these results are relatively vulnerable to unobserved bias. The robustness of the effects on general health indicators may be of interest in future research.

Reference

Adams, S. (2002). Educational attainment and health: Evidence from a sample of older adults. *Education Economics*, 10, 97-109.

Aizer, A., Eli, S., Ferrie, J., and Lleras-Muney, A. (2016). The Long-Run Impact of Cash Transfers to Poor Families, *American Economic Review*, 106(4), 935-71.

Battistin, E., and Sianesi, B. (2006). Misreported schooling and returns to education: evidence from the UK. *The Review of Economics and Statistics*, 93(2), 495–509.

Becker, G. S. (1993). The economic way of looking at behavior. *Journal of Political Economics*, 101(3), 385–409.

Becker, G. S., and Mulligan, C. B. (1997). The Endogenous Determination of Time Preference. *The Quarterly Journal of Economics*, 112(3), 729–758.

Black, D.A., and Smith, A.M. (2004). How robust is the evidence on the effects of college quality? Evidence from matching. *Journal of Econometrics*, 121 (1-2), 99-124.

Blundell, R., Dearden, L., and Sianesi, B. (2005). Evaluating the Impact of Education on Earnings in the UK: Models, Methods and Results from the NCDS. *Journal of the Royal Statistical Society: Series A,* 168(3), 473–512.

Braga, M., and Bratti, M. (2013). The effect of schooling on health: Evidence on several health outcomes and behaviours. Health, Econometrics and Data Group (HEDG) Working Papers 13/19.

Bynner, J., Elias, P., McKnight, A., Pan, H. and Pierre, G. (2002). Young People's Changing Routes to Independence. York: Joseph Rowntree Foundation.

Caliendo, M. and Kopeinig, S. (2008). Some Practical Guidance for the Implementation of Propensity Score Matching'. *Journal of Economic Surveys*, 22 (1), 31–72.

Case, A., and Deaton, A. (2005). Health and Wealth among the Poor: India and South Africa Compared. *American Economic Review*, 95(2), 229-233.

Chevalier, A. and Feinstein, L, (2007). Sheepskin or Prozac: The Causal Effect of Education on Mental Health. Working Papers 200715, Geary Institute, University College Dublin.

Clark, D., and Royer, H. (2013). The Effect of Education on Adult Mortality and Health:

Evidence from Britain. *The American Economic Review*, *103*(6), 2087–2120.

Conti, G., Heckman, J., and Urzua, S. (2010). The Education-Health Gradient. *American Economic Review,* 100(2)*,* 234–238

Contoyannis, P., Jones, A.M, and Rice, N. (2004). The dynamics of health in the British Household Panel Survey, *Journal of Applied Econometrics*, 19(4), 473-503.

Currie, J, and Moretti, E. (2003) Mother's Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings and Longitudinal Data *Quarterly Journal of Economics*, 118(4), 1495-1532.

Cutler, D., Deaton, A., and Lleras-Muney, A. (2006). The Determinants of Mortality. *Journal of Economic Perspectives*, 20(3), 97-120

Cutler, D., and Lleras-Muney. A. (2008). Education and Health: Evaluating Theories and Evidence, Making Americans Healthier: Social and Economic Policy as Health Polic*,* edited by Schoeni. R.F., House, J. S., Kaplan, G., and Pollack, H. New York: Russell Sage Foundation.

Cutler, D., and Lleras-Muney, A. (2010). Understanding differences in health behaviors by education. *Journal of Health Economics* 29, 1-28.

Dale, S. B., and Krueger, A. B. (2002). Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables. *The Quarterly Journal of Economics*, 117(4), 1491-1527.

Dale, S. B., & Krueger, A. B. (2014). Estimating the Effects of College Characteristics over the Career Using Administrative Earnings Data. *The Journal of Human Resources*, 49(2), 323-358.

Dehejia, R. H., and Wahba, S. (1999). Causal Effects in Non-experimental Studies: Reevaluating the Evaluation of Training Programmes. *Journal of American Statistical Association*, 94, 1053–1062.

de Luna, X and Lundin, M., (2014). Sensitivity analysis of the unconfoundedness assumption with an application to an evaluation of college choice effects on earnings. *Journal of Applied Statistics*, 41(8), 1767-1784.

Feinstein, L. (2002). Quantitative Estimates of the Social Benefits of Learning 2: Health (Depression and Obesity), Wider Benefits of Learning Research Report, 6.

Fuchs, V.R. (1982). Time Preference and Health: An Exploratory Study, Economic Aspects of Health.93-120. edited by V.R. Fuchs, Chicago: University of Chicago Press

Glied, A., and Lleras-Muney, A. (2003). Health Inequality, Education and Medical Innovation (June 2003). *NBER Working Paper* No. w9738.

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

Grossman, M. (2004) The Demand for Health, 30 Years Later: A Very Personal Retrospective and Prospective Reflection *Journal of Health Economics*, 23(4), 629-36.

Grossman, M. (2006). Education and Non-market Outcomes. *Handbook of the Economics of Education,* 1, 577-633.

Heckman, J., Ichimura, H., and Todd, P. (1997). Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. *Review of Economic Studies*, 64, 605–654.

Holland, P. (1986). Statistics and Causal Reasoning. *Journal of the American Statistical Association*. 81, 945-960.

Huisman, J.M., Matthijs, H.C.P., and Visser, P.M. (2005). Harmful Cyanobacteria. Springer Aquatic Ecology Series 3. Dordrecht.

Jürges, H., Reinhold, S., and Salm, M. (2011). Does schooling affect health behavior? Evidence from the educational expansion in Western Germany. *Economics of Education Review*, 30, 862-872.

Jürges, H., Kruk, E. and Reinhold, S. (2013). The effect of compulsory schooling on health evidence from biomarkers. *Journal of Population Economics*, 26, 645–672

Lechner, M. (2002). Program heterogeneity and propensity score matching: An application to the evaluation of active labor market policies. *The Review of Economics and Statistics* 84(2), 205–220.

Link, B.G., and Phelan J. (1995). Social conditions as fundamental causes of disease. *Journal Health and Social Behaviour*, Spec No, 80-94.

Lochner, L. (2011). Nonproduction Benefits of Education: Crime, Health, and Good Citizenship, Editor(s): Eric A. Hanushek, S., Machin, L., Woessmann, *Handbook of the Economics of Education*, 4.

Oreopoulos, P. (2006). Estimating Average and Local Average Treatment Effects of Education when Compulsory Schooling Laws Really Matter. *American Economic Review*, 96 (1), 152-175.

Rosenbaum, P. and Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41-55.

Rubin, D.B. (1974). Estimating Causal Effects of Treatments in Randomised and Non-randomised Studies, *Journal of Educational Psychology*, 66, 688-701.

Rubin, D.B. (1978). Bayesian Inference for Causal Effects: The Role of Randomization. *The Annals of Statistics*, 6(1), 34-58.

Rubin, D.B. (1997). Estimating Causal Effects from Large Data Sets Using the Propensity Score. Annals of Internal Medicine, 8(2), 757-763.

Sabates, R., and Feinstein, L. (2004) Education, training and the take-up of preventative health care, *Wider Benefits of Learning Research Report*, 12.

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

Saxton, J. (2000). Investment in education: Private and public returns. Retrieved from http://www.house.gov/jec/educ.pdf.

Schultz, T. W., 1961. Investment in Human Capital. *American Economic Review*, 51(1), 1–17.

Wolfe, B.L., and Zuvekas, S.H. (1997). Nonmarket Outcomes of Schooling. *International Journal of Educational Research, 27*, 491-501.

Sianesi, B. (2004). An Evaluation of the Swedish System of Active Labor Market Programs in the 1990s, *The Review of Economics and Statistics*, 86(1), 133-155,

Meghir, Costas, Palme, M., and Simeonova, E. (2018). Education and Mortality: Evidence from a Social Experiment. *American Economic Journal: Applied Economics*, 10 (2), 234-56.

Peel, M. and Makepeace, G. (2012). Differential audit quality, propensity score matching and Rosenbaum bounds for confounding variables. *Journal of Business Finance and Accounting*, 39 (5&6), 606–648.

Rosenzweig, M. R., & Schultz, T. P. (1983). Estimating a Household Production Function: Heterogeneity, the Demand for Health Inputs, and Their Effects on Birth Weight. *Journal of Political Economy*, 91(5), 723–746.

Rosenbaum, P. R. (2012) Testing one hypothesis twice in observational studies, *Biometrika*, 99(4), 763-774

Ross, C. E. and Wu, C. (1995) The links between education and health. *American Sociological Review*, 60(5), 719-45.

Russell, G., and Shaw, S. (2009). A study to investigate the prevalence of social anxiety in a sample of higher education students in the United Kingdom. *Journal of Mental Health,* 18(3), 198–206.

Voigtländer. N, Voth, H. (2012). Persecution Perpetuated: The Medieval Origins of Anti-Semitic Violence in Nazi Germany. *The Quarterly Journal of Economics*, 127(3), 1339-1392.

White, I., Blane, D., Morris, J. and Mourouga, P. (1999) Educational attainment, deprivation-affluence and self-reported health in Britain: a cross-sectional study. *Journal of Epidemiology & Community Health*, 53(9), 535-41.

Appendix A

Table A.1 Estimation of Probit model for Propensity score

Ethics (non-White)	0.025	**Number of siblings in 1974**	-0.054
Mathematics ability at 7 years		**Father's interest in education**	
5th quintile (highest)	0.194	Over concerned	0.596
4th quintile	-0.014	Very interested	0.181
3rd quintile	-0.156	Some interest	0.055
2nd quintile	-0.354	**Mother's interest in education**	
1st quintile (lowest)	-0.254	Over concerned	0.198
Reading ability at 7 years		Very interested	0.068
5th quintile (highest)	0.045	Some interest	0.007
4th quintile	0.022	**Bad finances in 1969 or 1974**	-0.112
3rd quintile	-0.263	**Region in 1974**	
2nd quintile	-0.209	North West	-0.175
1st quintile (lowest)	-0.204	North	0.198
Mathematics ability at 11 years		East and West	-0.008
5th quintile (highest)	-0.466	East	-0.448
4th quintile	-0.402	London and South East	0.249
3rd quintile	-0.171	South	-0.345
2nd quintile	-0.002	South West	0.037
1st quintile (lowest)	0.231	Midlands	-0.148
Reading ability at 11 years		Wales	omitted
5th quintile (highest)	0.159	Scotland	-0.560
4th quintile	0.068		
3rd quintile	-0.110		
2nd quintile	-0.298	**Father's years of education**	0.275
1st quintile (lowest)	-0.505	**Mother's years of education**	0.387
Comprehensive school 1974	0.193		
Secondary modern school 1974	0.112	**Birth weight (<2500g)**	0.086
Grammar school 1974	0.387	**Mother Smoking During Pregnancy**	
Private school 1974	0.560	Non smoker	0.002
Other school 1974	-0.017	Medium smoker	0.041
Father's social class in 1974		Heavy smoker	0.012
Professional	0.474	Variable smoker	0.058
Intermediate	0.234	**General health at age 7**	
Skilled Non-manual	0.182	Good	0.100
Skilled manual	-0.063	Abnormal	0.090
Semi-skilled non-manual	0.008	Missing Value	0.121
Semi-skilled manual	-0.025	**General health at age 11**	
Unskilled	-0.287	Good	0.045

Missing, or unemployed or no father	Omitted	Abnormal	0.025
		Missing Value	0.097
Behaviour Score 7		**General health at age 11**	
Normal	0.015	Good	0.067
Moderate problem	-0.012	Abnormal	0.045
Severe problem	-0.045	Missing Value	0.089
Missing or Incomplete	0.025	**Behaviour Score 16**	
Behaviour Score 11		Normal	0.045
Normal	0.044	moderate	-0.001
moderate	0.065	Severe problems	-0.094
Severe problems	-0.058	Missing or Incomplete	0.056
Missing or Incomplete	0.045		

Table A.2 Returns to HE on health and health-related conditions, PSM results, thick support region

		PSM NN			**PSM Kernel**	
	Full	**Male**	**Female**	**Full**	**Male**	**Female**
Panel A						
			SRH			
Age 33	-0.097*	-0.082*	-0.132**	-0.091*	-0.079*	-0.125**
Age 42	-0.106*	-0.115**	-0.153**	-0.101*	-0.109**	-0.151**
Age 50	-0.111*	-0.125*	-0.199**	-0.107*	-0.120*	-0.185**
			BMI			
Age 33	-0.356*	-0.400**	-0.213	-0.323*	-0.390**	-0.203
Age 42	-0.511**	-0.598**	0.412**	-0.505**	-0.590**	0.400**
Age 50	-0.679**	-0.899**	-0. 513**	-0.661**	-0.873**	-0. 507*
			Obesity			
Age 33	-0.036	-0.098	-0.026	-0.035	-0.097	-0.0234
Age 42	-0.156**	-0.175**	-0.154**	-0.149**	-0.169**	-0.150**
Age 50	-0.187**	-0.201**	-0.181**	-0.180**	-0.197**	-0.173**
Panel B						
			Alcohol Drinking Frequency			
Age 33	-0.270**	-0.113	-0.295**	-0.250**	0.107	-0.282**
Age 42	-0.352**	-0.200*	-0.470	-0.343**	-0.194*	-0.459
Age 50	-0.358**	-0.201*	-0.474**	-0.345**	-0.194*	-0.456**
			Smoke Frequency			
Age 33	-0.198**	-0.145	-0.281**	-0.189**	-0.143	-0.265**
Age 42	-0.135**	-0.083	-0.139**	-0.128**	-0.078	-0.135**
Age 50	-0.114	-0.069	-0.121	-0.110	-0.063	-0.116
Panel C						
			Depression			
Age 33	Full	Male	Female	Full	Male	Female
	-0.012	-0.003	-0.036	-0.010	-0.003	-0.034
Age 42	-0.017	-0.011	-0.082	-0.016	-0.010	-0.080

Age 50	-0.022	-0.017	-0.110	-0.015	-0.015	-0.106

Note **significant at the 5% level; *significant at 10% level

Table A.3 Joint quality of matching indicators

	PSM NN			PSM Kernel		
Panel A						
			SRH			
Before match						
	Full	**Male**	**Female**	**Full**	**Male**	**Female**
Mean absolute bias	13.82	12.44	18.55			
Pseudo R^2	0.11	0.10	0.19			
After match						
Age 33						
Mean absolute bias	4.84	4.56	6.27	3.75	3.96	5.15
Pseudo R^2	0.02	0.02	0.03	0.01	0.01	0.03
Age 42						
Mean absolute bias	3.65	4.74	5.96	3.57	4.29	5.57
Pseudo R^2	0.01	0.02	0.03	0.01	0.02	0.03
Age 50						
Mean absolute bias	3.22	3.45	3.25	2.45	2.84	2.67
Pseudo R^2	0.01	0.01	0.01	0.01	0.01	0.01
			BMI			
Before match						
	Full	Male	Female	Full	Male	Female
Mean absolute bias	26.12	27.45	30.12			
Pseudo R^2	0.31	0.29	0.34			
After match						
Age 33						
Mean absolute bias	7.84	6.56	6.27	5.49	5.79	5.87
Pseudo R^2	0.05	0.04	0.04	0.03	0.03	0.03
Age 42						
Mean absolute bias	6.29	7.14	7.45	6.22	6.80	6.45
Pseudo R^2	0.03	0.04	0.04	0.02	0.02	0.02
Age 50						
Mean absolute bias	7.46	8.01	7.12	6.45	5.78	5.49
Pseudo R^2	0.05	0.05	0.04	0.03	0.03	0.02
			Obesity			
Before match						
	Full	Male	Female	Full	Male	Female
Mean absolute bias	10.13	9.46	9.65			
Pseudo R^2	0.07	0.07	0.06			
After match						
Age 33						
Mean absolute bias	2.13	2.07	2.04	1.98	1.54	1.45
Pseudo R^2	0.005	0.005	0.005	0.00	0.00	0.00
Age 42						
Mean absolute bias	3.45	3.78	2.98	2.82	2.62	2.97
Pseudo R^2	0.01	0.01	0.01	0.00 5	0.005	0.005
Age 50						
Mean absolute bias	3.79	4.02	4.14	3.76	3.67	3.46
Pseudo R^2	0.01	0.02	0.02	0.01	0.01	0.01

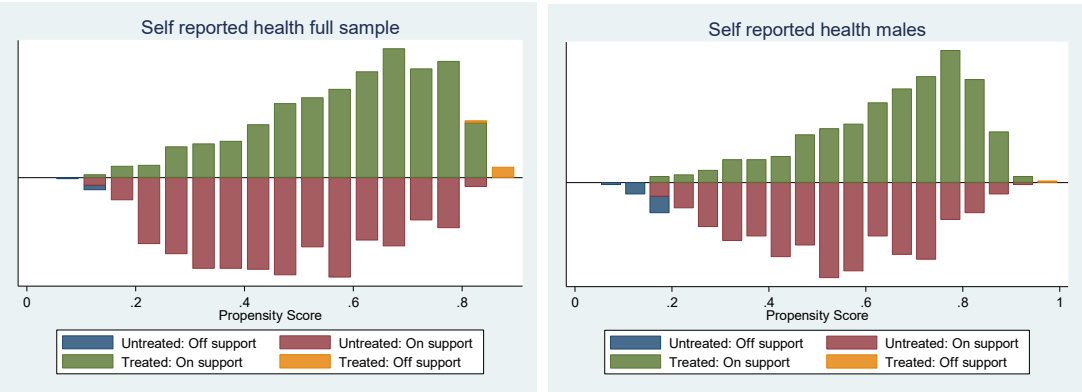
Panel B						
			Alcohol Drinking Frequency			
	Before match					
	Full	Male	Female	Full	Male	Female
Mean absolute bias	12.41	11.16	11.97			
Pseudo R^2	0.12	0.11	0.12			
After match						
Age 33						
Mean absolute bias	4.84	4.56	6.27	3.75	3.96	5.15
Pseudo R^2	0.02	0.02	0.03	0.01	0.01	0.03
Age 42						
Mean absolute bias	4.87	4.13	4.41	3.13	2.45	3.01
Pseudo R^2	0.02	0.02	0.02	0.01	0.01	0.01
Age 50						
Mean absolute bias	3.79	4.02	4.14	3.76	3.67	3.46
Pseudo R^2	0.01	0.02	0.02	0.01	0.01	0.01
			Smoke Frequency			
Before match						
	Full	Male	Female	Full	Male	Female
Mean absolute bias	14.17	13.64	13.75			
Pseudo R^2	0.12	0.11	0.12			
After match						
Age 33						
Mean absolute bias	8.41	8.45	8.17	6.13	6.47	6.97
Pseudo R^2	0.05	0.06	0.05	0.01	0.01	0.03
Age 42						
Mean absolute bias	7.16	7.57	7.13	7.01	6.13	6.48
Pseudo R^2	0.04	0.04	0.04	0.04	0.03	0.03
Age 50						
Mean absolute bias	3.79	4.02	4.14	3.76	3.67	3.46
Pseudo R^2	0.01	0.02	0.02	0.01	0.01	0.01
Panel C						
			Depression			
Before match						
	Full	Male	Female	Full	Male	Female
Mean absolute bias	12.16	11.71	12.23			
Pseudo R^2	0.11	0.10	0.11			
After match						
Age 33						
Mean absolute bias	6.18	5.39	6.40	5.56	4.79	5.57
Pseudo R^2	0.03	0.03	0.03	0.03	0.02	0.03
Age 42						
Mean absolute bias	7.16	7.57	7.13	7.01	6.13	6.48
Pseudo R^2	0.04	0.04	0.04	0.04	0.03	0.03
Age 50						
Mean absolute bias	7.64	7.24	7.61	7.12	6.97	7.00
Pseudo R^2	0.04	0.03	0.04	0.03	0.03	0.03

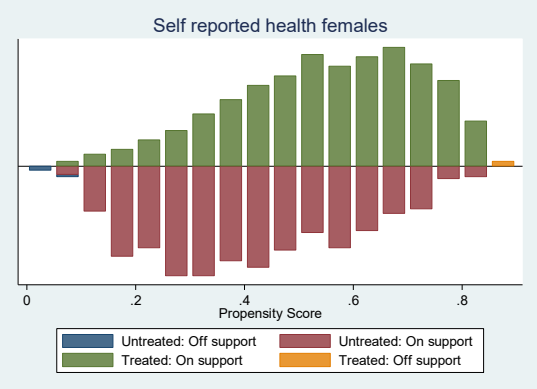
Table A.4 Rosenbaum Bounds for PSM estimation on different health outcomes

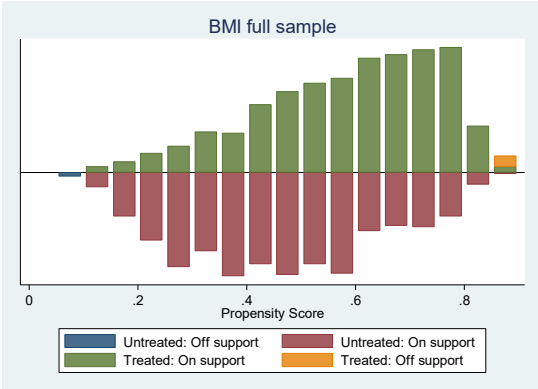
	PSM NN			PSM Kernel		
Panel A						
			SRH			
Age33						
Γ cut-off	1.15	1.16	1.14	1.16	1.17	1.16
P value	12.0%	11.6%	12.0%	11.5%	12.4%	10.7%
Age 42						
Γ cut-off	1.06	1.07	1.07	1.15	1.12	1.10
P value	9.6%	10.2%	11.1%	9.5%	11.2%	12.4%
Age 50						
Γ cut-off	1.13	1.15	1.12	1.11	1.10	1.11
P value	11.2%	12.4%	10.2%	10.2%	9.7%	11.3%
			BMI			
Age33						
Γ cut-off	1.24	1.20	1.23	1.25	1.26	1.25
P value	6.2%	6.4%	5.8%	5.4%	5.2%	5.1%
Age 42						
Γ cut-off	1.19	1.20	1.22	1.21	1.22	1.21
P value	5.5%	5.4%	5.8%	5.1%	5.3%	5.4%
Age 50						
Γ cut-off	1.17	1.18	1.17	1.20	1.21	1.24
P value	5.2%	5.4%	5.4%	5.2%	5.7%	5.4%
			Obesity			
Age33						
Γ cut-off	1.13	1.12	1.11	1.15	1.17	1.18
P value	8.2%	7.2%	8.4%	6.7%	5.8%	5.4%
Age 42						
Γ cut-off	1.13	1.13	1.12	1.12	1.11	1.12
P value	5.1%	6.3%	5.4%	5.2%	5.2%	5.3%
Age 50						
Γ cut-off	1.12	1.11	1.13	1.15	1.16	1.12
P value	5.2%	5.1%	5.2%	5.5%	5.8%	5.9%
Panel B						
			Alcohol Drinking Frequency			
Age33						
Γ cut-off	1.42	1.41	1.40	1.43	1.45	1.41
P value	6.2%	5.2%	5.4%	5.7%	5.8%	5.4%
Age 42						
Γ cut-off	1.40	1.43	1.42	1.44	1.45	1.44
P value	5.2%	5.6%	5.5%	5.2%	5.2%	5.8%
Age 50						

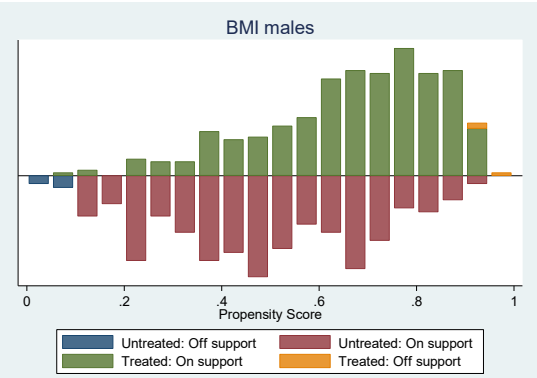
Γ cut-off	1.40	1.37	1.38	1.45	1.42	1.41
P value	5.5%	5.7%	5.8%	5.4%	5.8%	5.5%
			Smoke Frequency			
Age33						
Γ cut-off	1.32	1.31	1.37	1.38	137	1.39
P value	5.2%	5.2%	5.4%	5.7%	5.8%	5.4%
Age 42						
Γ cut-off	1.35	1.34	1.32	1.38	1.37	1.38
P value	5.1%	5.2%	5.4%	5.2%	5.2%	5.1%
Age 50						
Γ cut-off	1.35	1.35	1.35	1.40	1.40	1.41
P value	5.1%	5.1%	5.2%	5.4%	5.3%	5.2%
Panel C						
			Depression			
Age33						
Γ cut-off	1.54	1.57	1.58	1.62	1.64	1.61
P value	5.2%	5.2%	5.4%	5.1%	5.3%	5.4%
Age 42						
Γ cut-off	1.61	1.59	1.54	1.65	1.67	1.68
P value	5.1%	5.2%	5.4%	5.2%	5.2%	5.1%
Age 50						
Γ cut-off	1.55	1.58	1.61	1.65	1.66	1.68
P value	5.1%	5.1%	5.2%	5.4%	5.3%	5.2%

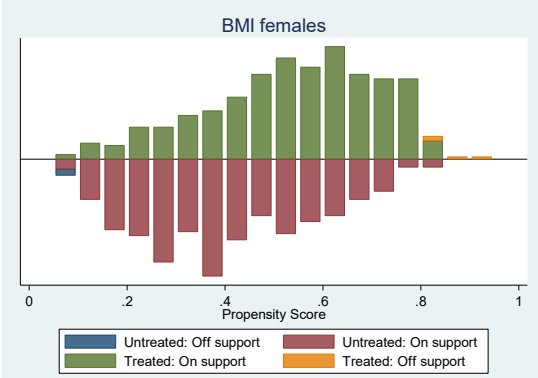
Figure A.1 Propensity score distributions and common support regions

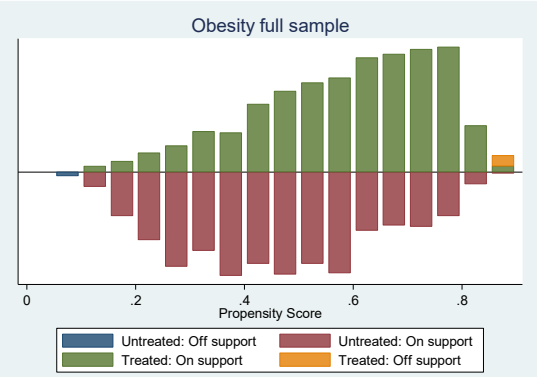


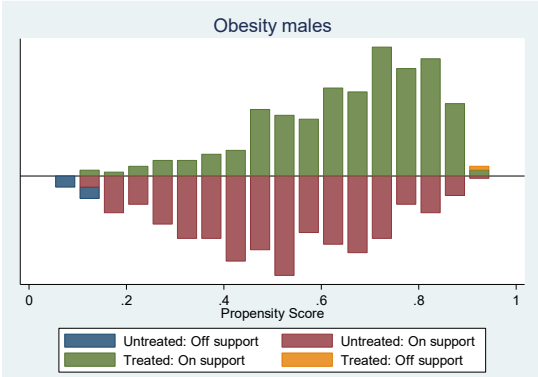

Self reported health females
0
.2
.4
.6
.8
Propensity Score
Untreated: Off support
Untreated: On support
Treated: On support
Treated: Off support

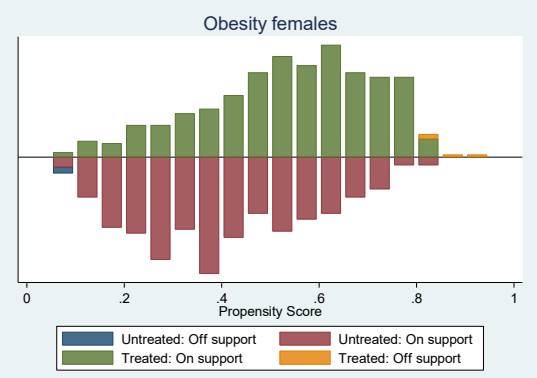

BMI full sample
0
.2
.4
.6
.8
Propensity Score
Untreated: Off support
Untreated: On support
Treated: On support
Treated: Off support

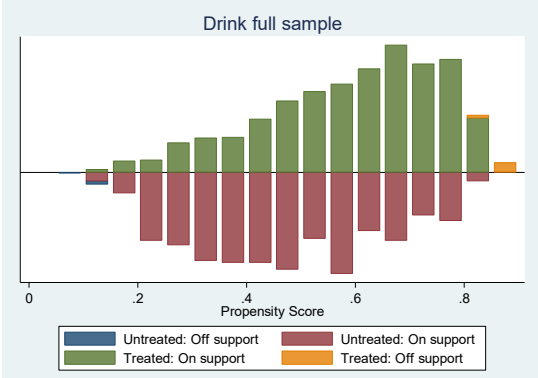

BMI males
0
.2
.4
.6
.8
1
Propensity Score
Untreated: Off support
Untreated: On support
Treated: On support
Treated: Off support

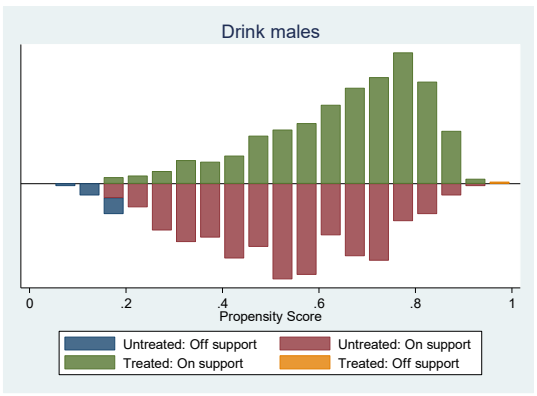

BMI females
0
.2
.4
.6
.8
1
Propensity Score
Untreated: Off support
Untreated: On support
Treated: On support
Treated: Off support

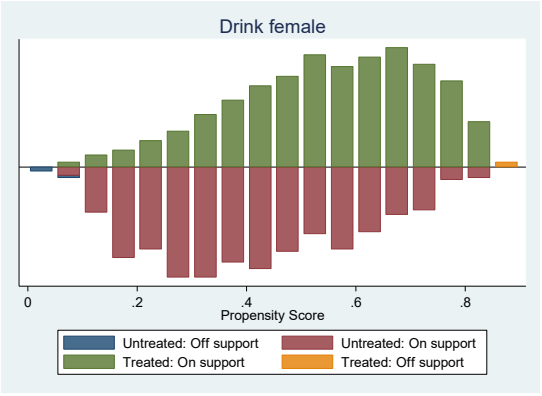

Obesity full sample
0
.2
.4
.6
.8
Propensity Score
Untreated: Off support
Untreated: On support
Treated: On support
Treated: Off support

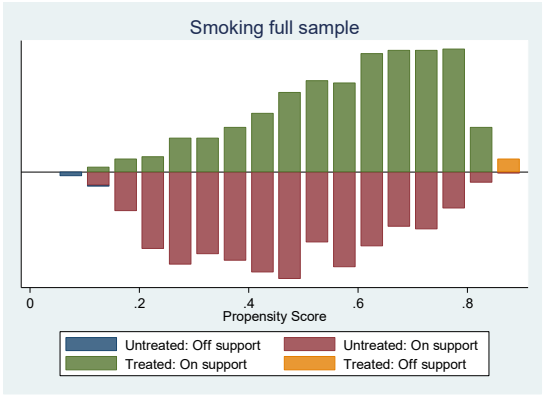

Obesity males
0
.2
.4
.6
.8
1
Propensity Score
Untreated: Off support
Untreated: On support
Treated: On support
Treated: Off support

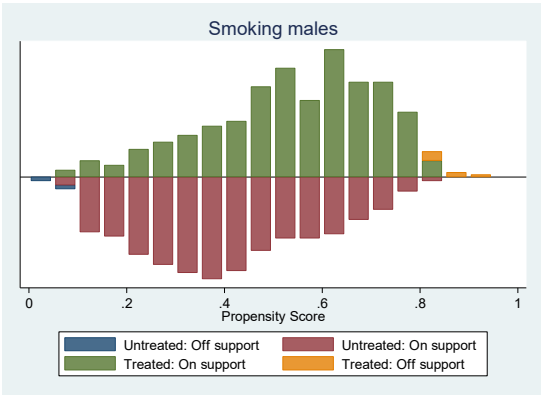

Obesity females
0
.2
.4
.6
.8
1
Propensity Score
Untreated: Off support
Untreated: On support
Treated: On support
Treated: Off support

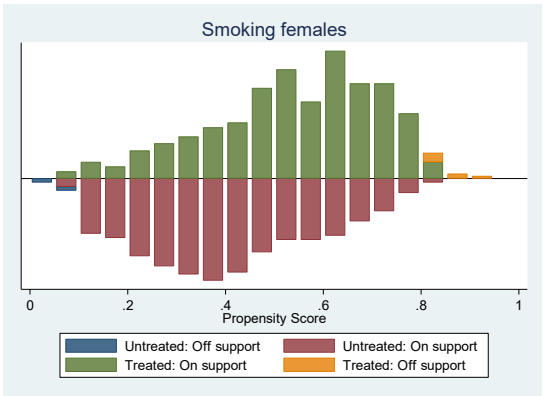

Drink full sample
0
.2
.4
.6
.8
Propensity Score
Untreated: Off support
Untreated: On support
Treated: On support
Treated: Off support

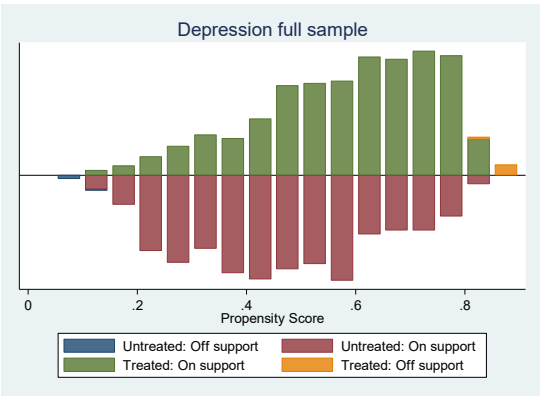

Drink males
0
.2
.4
.6
.8
1
Propensity Score
Untreated: Off support
Untreated: On support
Treated: On support
Treated: Off support

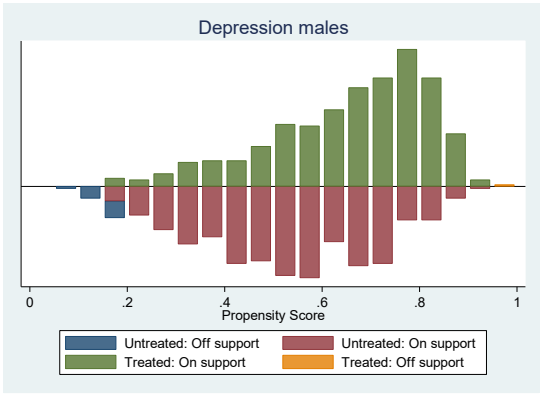

Drink female
0
.2
.4
.6
.8
Propensity Score
Untreated: Off support
Untreated: On support
Treated: On support
Treated: Off support

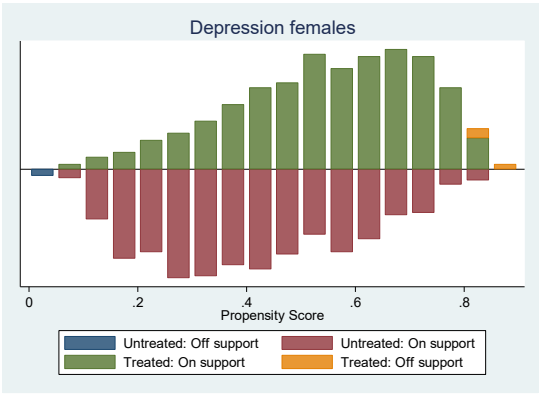

Smoking full sample
0
.2
.4
.6
.8
Propensity Score
Untreated: Off support
Untreated: On support
Treated: On support
Treated: Off support


Smoking males
0
.2
.4
.6
.8
1
Propensity Score
Untreated: Off support
Untreated: On support
Treated: On support
Treated: Off support


Smoking females
0
.2
.4
.6
.8
1
Propensity Score
Untreated: Off support
Untreated: On support
Treated: On support
Treated: Off support


Depression full sample
0
.2
.4
.6
.8
Propensity Score
Untreated: Off support
Untreated: On support
Treated: On support
Treated: Off support


Depression males
0
.2
.4
.6
.8
1
Propensity Score
Untreated: Off support
Untreated: On support
Treated: On support
Treated: Off support


Depression females
0
.2
.4
.6
.8
Propensity Score
Untreated: Off support
Untreated: On support
Treated: On support
Treated: Off support

Appendix B

(a) Nearest neighbour matching

One of the most straightforward methods to implement matching estimators is Nearest Neighbour (NN) matching (Caliendo and Kopeinig, 2008). In its simplest pairwise matching, it starts from each treated unit's propensity score and tries to find a control unit with the closest or most similar estimated propensity scores to use as a match. Once each treated unit is matched with a control unit, the difference between the outcome of the treated units Y_i^T and the outcome of the matched untreated units Y_j^C is computed.

The outcome of treated unit *i* is matched to a control unit *j* with the closest propensity score: $C(i) = min||p(X_i) - p(X_j)||$. The ATT is then obtained by averaging these differences[11] given by:

$$ATT = \frac{1}{N_T}\sum_{i\in T}\{w_{1i} - w_{0j}\} \quad (22)$$

This method is usually applied *with replacement*, in the sense that an untreated unit can be the best match for more than one treated unit. Each treated unit can only be used once, but the same control unit may be used more than once if it is the closest match for many different treatment units.

(b) Kernel matching

Kernel matching is a non-parametric matching estimator which uses weighted averages of all individuals in the control group within the common support region to construct the counterfactual outcome. In kernel matching, the outcome of the treated unit *i* is matched to a weighted average of the outcomes of possibly all control units and the weight is set to:

$$g_{ij} = \frac{K\left(\frac{p_i - p_j}{h}\right)}{\sum K\left(\frac{p_i - p_j}{h}\right)} \quad (23)$$

Where $K(.)$ denotes a non-negative and symmetric kernel function and h denotes the bandwidth. ATT advanced by kernel matching is given by:

$$ATT = \frac{1}{N_T}\sum_{i\in T}\{w_{1i} - \sum_{j\in C(i)} g_{ij} w_{0j}\} \quad (24)$$

The advantage of the Kernal matching approach is that a lower variance can be achieved since more information on control groups is used. However, the estimated results are often very

[11] The derivation of formula and variance formula can be seen in See Becher and Ichino (2002).

sensitive to the choice of bandwidth. A high value of bandwidth parameter yields a smoother estimated density function, a better fit and a lower variance between the estimated and the true underlying density function, nonetheless also leading to a possible biased estimate.