Education, dietary intakes and exercise

Stephanie von Hinke

Discussion Paper 21/748

23 June 2021

School of Economics

University of Bristol
Priory Road Complex
Bristol
BS8 1TU
United Kingdom

University of BRISTOL
School of Economics

Education, dietary intakes and exercise

Stephanie von Hinke *

June 23, 2021

Abstract

This paper examines the relationship between education and health behaviours, focusing on potential offsetting responses between calories *in* (i.e. dietary intakes) and calories *out* (i.e. physical activity). It exploits the 1972 British compulsory schooling law that raised the minimum school leaving age from 15 to 16 to estimate the effects of education on diet and exercise around middle age. Using a regression discontinuity design, the findings suggest that the reform led to a *worsening* of the quality of the diet, with increases in total calories, fats and animal proteins. However, I find that these changes are partially offset by a discontinuous increase in physical activity. Back-of-the-envelope calculations suggest little effect on the balance of calories. As such, the findings show that focusing on the two components of energy balance provides additional information that is concealed in analyses that only use a measure of obesity.

Keywords: Compulsory schooling; Offsetting responses; Nutritional choices; Physical activity; Instrumental Variables

JEL classification: I12, I21, I28

No. of words: Approx 9,500 (excluding tables and figures, but including appendices)

Acknowledgements: I would like to thank John Cawley, Matt Dickson, Eleonora Fichera, Katharina Janke, Hans van Kippersluis, Birgitta Rabe, Hans Sievertsen, seminar participants at Erasmus University Rotterdam, the Toulouse School of Economics, and participants at the Royal Economic Society and the Health Economics Study Group conference for comments on a previous version of the paper. I gratefully acknowledge financial support from the UK Medical Research Council (G1002345).

*__Correspondence:__ Stephanie von Hinke, University of Bristol, School of Economics, 12 Priory Road, Bristol BS8 1TU, S.vonHinke@bristol.ac.uk; Erasmus University Rotterdam, Erasmus School of Economics; Institute for Fiscal Studies.

1 Introduction

There is a strong social gradient in health and disease, with lower socio-economic classes experiencing increased morbidity. One of the main causes of this are poor dietary choices (Marmot, 2005; WHO, 2002). Governments across the world are trying to encourage individuals to make healthier choices, through channels such as information provision (e.g. the five-a-day campaign (Capacci and Mazzochi, 2011), food labelling (Fichera and von Hinke, 2020)), and fiscal measures (e.g. taxes (Fletcher, 2010), targeted benefits (Griffith et al., 2018)).

Given the social gradient in nutritional choices, one question is whether we can improve individuals' diets by increasing their socio-economic status, such as education. Indeed, the model for the demand for health (Grossman, 1972) suggests that education affects health *directly* via the accumulation of knowledge and improved cognitive functioning. For example, education can change individuals' allocative efficiency, affecting the allocation of health inputs such as dietary choices. Education may also *indirectly* affect health inputs by increasing earnings, which can in turn affect nutritional choices. For example, higher wages may increase the affordability of health-improving foods. However, higher wages also increase the opportunity cost of time, potentially leading to individuals increasing their consumption of (time-saving) ready meals, which tend to be less healthy.[1]

This paper examines the relationship between education and the nutritional composition of the diet, with two main contributions. First, with little evidence on the causal effect of education on dietary choices, this paper fills this gap in the literature. Because of the importance of dietary choices in determining individuals' body weight, one could indirectly explore whether education affects diets by investigating its effect on obesity. Indeed, various studies have taken this approach, but they tend to find no consistent evidence of such a relationship (for a recent review, see Galama et al. (2018)). However, as obesity is a function of calories *in* as well as *out*, using obesity as the main outcome may conceal more disaggregated effects coming from the two separate components.

The second contribution, therefore, comes from being able to shed light on the potential offsetting responses between calories in and calories out. Hence, I investigate the causal effect of education on dietary choices as well as individuals' physical activity. I exploit national changes to the UK minimum school leaving age introduced in 1972 as exogenous variation in years of schooling in an IV setup. More specifically, the UK government increased the age at which individuals were legally allowed to leave school from 15 to 16 on 1 September 1972. This meant that everyone born prior to 1 September 1957 was allowed to leave school when they turned 15, whereas those born on or after 1 September 1957 had to stay in school until they turned 16. The 1972 reform has been exploited for analyses in other contexts, including those estimating the effect of education on e.g. wages (see e.g. Harmon and Walker, 1995; Oreopoulos, 2006), health and mortality (see e.g. Clark and Royer, 2013; van Kippersluis et al., 2009; Davies et al.,

[1] Other indirect effects could run through education affecting individuals' time preferences, making them more future oriented (Fuchs (1982)), which in turn can lead to better health outcomes and behaviours. Similarly, higher educated individuals may respond differently to stress or anxiety (e.g. comfort eating versus exercising), or have systematically different peers and therefore be exposed to different behaviours and habits.

2018; Janke et al., 2018), family formation (Geruso and Royer, 2018; Hener and Wilson, 2015), teenage motherhood (Wilson, 2014) and decision making (see e.g. Banks et al., 2018).[2]

Consistent with these studies, I find that the 1972 schooling reform increased years of education. The results for dietary choices, however, suggest that additional education *reduces* the nutritional quality of the diet. More specifically, I find that education increases purchases of total calories, fats and animal proteins. These findings are robust to a range of alternative model specifications, and I replicate these results using a different dataset to look at nutrient *consumption* as opposed to nutrient *purchases*. The results show similar patterns, though with a substantially smaller sample size, they are not significantly different from zero.

Next, I explore potential mechanisms of this effect, highlighting two channels in particular. First, I explore the role of income in explaining the increase in nutrient purchases, but I find no evidence of income driving this effect. Second, I investigate the role of alcohol and smoking. The analysis shows some suggestive evidence that education increases alcohol (but not tobacco) spending. Since alcohol and food are often considered complements, this suggests that the increased alcohol spending among the higher educated may explain some of the increase in nutrient purchases among this group.

Finally, I investigate the effect of education on calories *out*, exploring whether the worsening of the diet is mitigated by changes in levels of (sports-related) physical activity. Using the same empirical strategy, I find that additional education *increases* physical activity, suggesting that the reduction in the dietary quality is partially offset by increased physical exercise. Furthermore, I find no significant change in occupation-related strenuousness. A back-of-the-envelope calculation suggests that the change in calories *in* is likely to be only slightly larger compared to the change in calories *out*. Although these are rather crude estimates due to data limitations and measurement error, this suggests that there is little effect on the balance of calories. There are at least two explanations for this finding. First, individuals may choose to exercise more to allow them to consume more unhealthy foods (i.e., a licensing effect), and this may differ by education. Second, higher educated individuals may be more aware of the importance and benefits of physical activity (i.e., a knowledge effect), and doing more exercise requires more foods to sustain this higher level of physical activity. Although I am unable to identify the exact mechanism (e.g., a licensing versus knowledge effect), the analysis highlights the importance of separately exploring the effects on calories in versus calories out, providing additional information that is concealed in the analyses that only use measures of obesity.

With that, this paper speaks to a large literature that explores the effects of education on health. Much of this literature exploits compulsory schooling reforms to estimate the effects, with some mixed results. Whilst many studies suggest education does not have strong causal effects on morbidity or mortality (see e.g. Clark and Royer, 2013; Albouy and Lequien, 2009; Jurges et al., 2013, the latter possibly driven by a lack of power), some do find evidence of a negative effect (e.g. Lleras-Muney, 2005; van Kippersluis et al., 2009; Davies et al., 2018), or find such evidence for a selection of health outcomes (e.g. Janke et al., 2018). Other effects of

[2]Other approaches that have been used to explore the effect of education include Randomized Controlled Trials and twin studies (for a review of this literature, see Galama et al., 2018).

compulsory schooling that have been found in the literature are improvements in self-assessed health (Silles, 2009), reductions in long-term illness (Kemptner et al., 2011), worse mental health (potentially because it forces low achieving teenagers to remain in an academic environment; Avendano et al., 2017), but no changes in health knowledge (Johnston et al., 2015), smoking or obesity (see e.g. Galama et al., 2018). To the best of my knowledge, this study is the first to include a detailed investigation of the effect of education on the two main subcomponents that determine obesity, calories in and calories out, and with that show their substantial offsetting effects.

A handful of studies have explored the specific effects on nutritional choices. For example, Li and Powdthavee (2015) estimate the effect of education on the likelihood of complying with certain dietary guidelines (e.g. regular consumption of fruit and vegetables). They find that education increases compliance, with some improvements in self-assessed health. Atella and Kopinska (2014) and Fletcher (2015) estimate the effect of education on total calorie intakes, but not nutrients, whilst Barcellos et al. (2017) focus on macronutrients, as measured from respondents' reported consumption of 21 specific food groups. Using a similar research design that exploits compulsory schooling laws in the United States, Fletcher (2015) finds that an additional year of schooling decreases daily caloric intake by 87 calories and increases vitamin use (another dimension of dietary quality) by 31 percentage points, though these are not statistically significantly different from zero. The findings in this paper suggest the opposite for the UK: a drop in dietary quality, albeit offset by an increase in physical activity.

The paper also contributes to the literature that examines individuals' compensatory responses. The existing evidence on such effects within economics includes, for example, evidence on how individuals offset additional calories consumed in restaurants by cutting back on caloric intakes at other times (Anderson and Matsa, 2011); how parental investments in children may compensate for adverse (early life) shocks (see e.g. Almond and Currie, 2011); or how tobacco tax hikes lead smokers to compensate by extracting more nicotine per cigarette (see e.g. Adda and Cornaglia, 2006). The literature that specifically explores interactions between energy intakes and expenditures discusses homeostatic body weight regulation, which implies that body weight returns to base-line after any impositions, due to (innate) compensatory adjustments in energy balance (see e.g. Epstein and Wing, 1980; King et al., 2007). Similarly, the economics literature emphasizes the importance of both channels to understand recent changes in obesity over time (Cutler et al., 2003; Griffith et al., 2016). This paper confirms this compensatory channel and shows that the reduction in dietary quality due to additional education is offset by increased physical exercise, or vice versa, that the increased physical activity due to additional education is offset by increased food intake. Taken together, there is little effect on the balance of calories consumed.

The rest of the paper is structured as follows: the next section explains the compulsory schooling reform. Section 3 describes the data, followed by the empirical approach in Section 4. Section 5 presents the results, and Section 6 concludes.

2 The compulsory schooling reform

On 1 September 1972, the UK increased the age at which individuals were legally allowed to leave school from 15 to 16. This meant that everyone born prior to 1 September 1957 was allowed to leave school when they turned 15, whereas those born on or after 1 September 1957 had to stay in school until they turned 16. Figure 1 illustrates the impact of this reform on the proportion of individuals leaving school at age ≥15, ≥16, and ≥17 years old. The horizontal axis shows the year of birth, with the vertical bar denoting 1958: the first full birth cohort subject to the 1972 compulsory schooling reform. The lines are obtained from locally weighted regressions, estimated separately for cohorts born before and after 1957. It shows a clear discontinuity in the probability of leaving school at age 16 or older for the cohorts born after this cut-point, with no large changes in the proportion of individuals leaving education at age 15 or 17.

[Figure 1 about here]

3 Data

3.1 The Living Cost and Food Survey

The main dataset I use is the 2003 to 2015 Living Cost and Food Survey (LCFS; known as the Expenditure and Food Survey (EFS) between 2001 and 2007), which provides information about spending patterns, cost of living, food purchases and nutrition. The data are repeated cross-sections of around 6,000 randomly selected households each year in the United Kingdom. Individuals are asked to keep a diary, where they record all daily expenditures for a period of 14 days, including foods purchased for home as well as out. Hence, the analysis below concerns food *purchases* as opposed to food *consumption*. The analysis focuses on the main shopper in the household, defined as the person with the highest food spending, who is assumed to be the main meal planner.

The advantages of these data are that they include a very detailed breakdown of households' shopping baskets, and of the nutrients that make up the shopping basket, allowing me to explore the effects of education on individual nutrients, such as unsaturated versus saturated fats, and animal versus plant protein.[3] Other data do not record nutritional information with this level of detail. Furthermore, surveys of food *consumption* that do include detailed nutritional information, such as the UK National Diet and Nutrition Survey (NDNS), generally contain much smaller sample sizes. Nevertheless, I will use these data to explore the robustness of my analyses below. As with all data, however, there are some disadvantages. One limitation is that the actual nutrient purchases are only measured at the level of the household. Hence, in

[3] These are high quality data, being used by the UK Government to define the 'basket of goods' for the Consumer Price Index (CPI) and the Retail Price Index (RPI), the former of which is used to uprate benefits, tax credits and public service pensions. The data are also used by HM Treasury and HM Revenues and Customs to study how taxes and benefits affect household income and to study the effects of policy in these areas (ONS, 2012).

the analyses, I relate the main shopper's level of education to *household* nutrient purchases. I show below, however, that the main shopper is responsible for the majority (83%) of all grocery spending.

Another limitation is that, similar to e.g. Oreopoulos (2006), I only observe year of birth, rather than month of birth. Hence, as the reform differentially affected those born before and after 1 September 1957, this does not allow me to take that into account in the analyses, meaning I cannot assign treatment status at the month-level. To avoid erroneously assigning individuals to being born before or after the cut-off, I drop the year 1957 from the analyses and compare those born up to 1956 to those born in 1958 onwards. This is also known as a "donut" regression discontinuity (Barreca et al., 2011). The advantage is that one does not need to worry about measurement error due to partially unobserved (or misreporting) of the running variable. However, a disadvantage is that the estimates require more extrapolation due to the omission of data immediately around the threshold. Furthermore, it means that treatment and control groups are not precisely assigned at the threshold and it does not allow me to capture local trends immediately around the cut-off using e.g. a local linear specification (see e.g. Clark and Royer, 2013).

To measure the quantity of nutrients purchased by each household, I use the conversion factors provided in the data. These measure the amount of each nutrient per kg of a highly disaggregated group of foods. Hence, multiplying the quantity purchased with its conversion factor gives the total amount of nutrients purchased over the 14 day period. I then divide this by 14 to obtain daily nutrient purchases. Finally, I create a 'nutrient equivalence scale' and use this to obtain the *average daily nutrient purchases per person.*[4]

I am interested in the following key nutrients that have been identified as important for health and development: energy (kcal), carbohydrates (including sugar and fibre), fats, and proteins. In addition, in line with current dietary guidelines, I will separately explore fats and saturated fats, and distinguish between total sugars and non-milk extrinsic sugars (NMES).[5] I will also separate out animal and plant protein. In addition to examining households' purchases of specific nutrients, I combine the nutritional information to obtain a score that indicates the 'healthiness' of the household diet. For this, I use the Nutritional Profile Model, developed by the UK Food Standards Agency (Rayner et al., 2009) and used by the UK Media regulator Ofcom. In short, this model attaches a score depending on the amount of each of the nutrients it contains per 100g. The Nutritional Profile Score measures the nutritional quality of the shopping basket, where higher scores indicate less healthy baskets. For more information on the Nutritional Profile Model, see Appendix A.

[4]This 'nutrient equivalence scale' takes into account the recommended daily energy intakes by age and gender, based on the Committee on Medical Aspects of Food Policy (COMA), published by the Department of Health (1991). More specifically, for each individual in the household, I take their (age and gender-specific) recommended energy intake as a proportion of the adult equivalent (i.e. 1900 kcal for 50-59 year old women, and 2550 kcal for 50-59 year old men) and sum this up over all household members. I use this to scale the total household nutrient purchases to individual purchases. However, the results are robust to using other equivalence scales, such as the OECD equivalence scales, the OECD modified equivalence scales, not using any equivalence scales, or simply dividing the amount of nutrients by the total number of household members.

[5]These are also known as 'added sugars' or 'free sugars', and include only those added to food and drinks.

To obtain the estimation sample, I drop households where the woman is pregnant, as this may affect the quantity and types of nutrients purchased. This leaves me with 74,080 households observed between 2003 and 2015. Next, I restrict the sample to the 1934-1982 birth cohorts and exclude 1957; i.e. 24 cohorts born prior to, and after the reform, leading to 60,764 households.[6] I check the robustness of my results by reducing this bandwidth below. Finally, I drop 5,178 households where the school leaving age is unknown, and drop five households who do not spend any money on food in the relevant 14-day period in which they were asked to keep a diary, leading to an estimation sample of 55,581 households.

3.2 Descriptive statistics

Table 1 presents the descriptive statistics for the final sample of households, where any individual-level characteristics refer to the main shopper. This shows that the main shopper is male in 34% of households, with an average age of 48. Across all households, the main shopper is responsible for 83% of all grocery spending.[7] Over half of the sample consists of households with two adult members, 15% have one child, and 20% have two or more children.

Figure C.1 and Figure C.2 in Appendix C show the distributions of the different nutrients, where all nutrients have been equivalised using the 'nutrient equivalence scale' discussed above, showing a near-normal distribution for the overall Nutritional Profile Score, and smooth densities for all individual nutrients.

[Table 1 about here]

4 Empirical methodology

I exploit the 1972 UK compulsory schooling reform as an instrument for education, using a regression discontinuity (RD) design. As not all individuals comply with the reform, I specify a fuzzy RD approach, estimated by two-stage least squares. This approach exploits the fact that individuals' years of schooling is a non-continuous function of the forcing variable (i.e. year of birth), with a discontinuity at the threshold (see e.g. Imbens and Lemieux, 2008; Lee and Lemieux, 2010). In other words, only an individual's year of birth determines whether or not the person is exposed to the schooling reform.

In the first stage, I specify a dummy D_i that equals 1 if the individual i is exposed to the reform (i.e. born in or after 1958) and zero otherwise as an instrument for the measure of education. I estimate the effects of this reform on a binary variable indicating whether the main shopper left school at age 16 or older, denoted by E_i:

$$E_i = \alpha_0 + \alpha_1 D_i + f(YOB_i) + \gamma \mathbf{X_i} + v_i,$$

[6] This cut-off ensures that individuals in the sample were not differentially affected by the 1947 compulsory schooling reform, as this affected those born from 1 April 1933 onwards.

[7] Distinguishing between household size, this is 100% for single-person households, 81% for two-person households, 76% for three-person households and 75% for four or more person households.

where $f(YOB_i)$ incorporates a non-linear (quadratic) function in the year of birth of the main shopper relative to the cut-off, as well as that interacted with the treatment dummy, allowing for differential non-linear trends for cohorts born before and after the threshold.[8] The vector $\mathbf{X_i}$ captures a set of background characteristics, including a quadratic in age, year and month dummies, gender, marital status, variables indicating the number of adults in the household, the number of children aged 0, aged 1, ..., aged 17, and region dummies. However, as these covariates should be uncorrelated to being born before or after the cut-off, its inclusion should not result in large changes in the estimate of the interest. Figure 1 graphically shows the extent to which the reform changed levels of education for the affected cohorts. The regression will quantify this shift and show the robustness to the inclusion of additional trends and covariates.

In the second stage, I examine the causal effect of education on nutritional purchases, using the dummy for being exposed to the reform (D_i) as the instrument for education (E_i) in a two-stage least squares (2SLS) regression.[9] The equation is given by:

$$Y_i^j = \beta_0 + \beta_1 E_i + f(YOB_i) + \zeta \mathbf{X_i} + \epsilon_i, \tag{1}$$

where Y_i^j denotes the average daily nutrient j purchased by individual i, with $i = 1, \ldots, N$ and $j = 1, \ldots, J$, and E_i denotes the measure of education that is instrumented by the dummy D_i. The function $f(YOB_i)$ again denotes the flexible function in the year of birth relative to the cut-off, $\mathbf{X_i}$ is a vector of covariates, and the error term is given by ϵ_i. All analyses are clustered at the birth cohort level. For all analyses, I use the weights provided by the data to adjust for non-response and ensure representativeness of the population.

It is worth highlighting some potential issues when exploiting the increase in the minimum school leaving age, which may affect the estimation and interpretation of the parameters. First, this policy may have led to complex general equilibrium effects, some of which may be working through the marriage market (see e.g. Akresh et al., 2018; Anderberg et al., 2019). For example, the rise in the minimum school leaving age may have resulted in more marriages *within* cohorts due to assortative mating, where the reform has increased education for both partners. In the analysis above, I only account for the main shopper's level of education, ignoring the spouse's education that may have similarly increased due to the reform, since spouse's education is missing for a large number of observations. This in turn may lead to an upward bias in the estimates. I present the analysis that accounts for spouse's education in the robustness section in Appendix B. Similarly, the rise in the minimum school leaving age may have affected marital sorting patterns. Indeed, Anderberg et al. (2019) find that the reform led to increases in never-married rates among the least educationally qualified in society. This differential sorting may

[8] I use a quadratic polynomial approach, rather than the global high-order polynomial approximation discussed in Gelman and Imbens (2019).

[9] As the different dependent variables sum up to the total household nutritional purchases, the errors are likely to be correlated across equations for a given household. In principle, this cross-equation correlation of the errors can be exploited to improve the efficiency of the estimates, using a three-stage least squares (3SLS) procedure. However, as the equations are just-identified, the 3SLS reduces to a 2SLS estimation, presented below.

independently affect individuals' decisions, including their nutritional choices.

Another potential issues is that, due to the higher wages driven by the increased education, higher educated individuals may decide to purchase *more expensive* or *higher quality* foods. As the analysis does not explore individuals' food expenditures itself, but instead focuses on total nutrients purchased, this is not an issue *per se*. However, it does affect the interpretation of the estimates. In particular, as discussed in Section 3, the dependent variable is obtained by multiplying the conversion factors for each highly disaggregated group of foods with the quantity purchased of that food group. The conversion factors record the *average* nutrients per kg of each food group. This implies that any systematic differences in the nutritional composition of *higher* compared to *lower* quality foods *within a food group* may bias the estimates. For example, if more expensive foods within a particular food group are nutritionally better (e.g. less sugar) and these are systematically more likely to be purchased by higher educated individuals, this would overestimate the parameter of interest.

Unfortunately, as the data provide nutrient conversion factors that differ only *across* food groups, but not *within*, it is not possible to explore the importance of this in more detail. However, the data include over 500 food groups in total, and although there is nutritional variation within these, they are designed to capture similar types of foods within each category (e.g. there are separate categories for oatmeal, muesli, high fibre breakfast cereals, sweetened breakfast cereals, other breakfast cereals, cereal bars, and cereal snacks). Other categories also distinguish between price/quality of foods (e.g. white bread premium, white bread soft grain, or white bread standard; less expensive beef steak, and more expensive beef steak), suggesting that the disaggregated food groups capture the majority of the variation in nutritional composition. Nevertheless, there remains some variation in the nutritional composition of foods *within* food groups (Griffith and O'Connell, 2009), suggesting that the estimates are likely to be an upper bound of the 'true' effect, which should be taken into account when interpreting the results.

Furthermore, because the parameters are estimated using IV, the estimates pick up a Local Average Treatment Effect (LATE). In other words, they capture the effect of staying in school until age 16 on nutritional choices for those who would have left at age 15 in the absence of the policy. This is a specific group of individuals, meaning that the estimates do not necessarily generalise to the population. Nevertheless, it is an interesting group for at least two reasons. First, as I show in Section 5 below, the reform affects a relatively large part of the population. And second, individuals affected by the policy are currently in their 60s, who have obesity rates that are among the highest in the UK (Baker, 2018).

5 Results

5.1 Nutritional choices

I start by discussing the findings from the naïve OLS regressions, followed by the IV analyses. Table 2 shows the former, regressing the Nutrient Profile Score (column 1) or the average amount of nutrients purchased per day (columns 2-13) on a binary indicator for whether the

individual left full-time education at age ≥16 and the other covariates discussed above. The results show that the higher educated, on average, have healthier diets: their Nutrient Profile Score is 0.25 units lower than the lower educated. With a standard deviation of the Nutrient Profile Score of 1.9 for cohorts born just before the cut-off (not shown here), this change is similar to approximately 13% of a standard deviation. Furthermore, the estimates suggest that the higher educated purchase significantly fewer calories, carbohydrates (including NMES), starch, fats (including saturated fats), sodium, and animal protein, whilst they purchase more fibre and vegetable protein. For example, households where the main shopper left full-time education at age ≥16 purchase 51 fewer calories per person per day. With average purchases of 2610 calories per person for cohorts born just before the cut-off, this is approximately 2%. For the other nutrients, the differences between nutrient purchases of those who left education at age ≥16 and those who left before are all in the range of 1-6%, apart from proteins, where it is less than 1%. However, the latter is driven by a positive correlation with vegetable protein and a negative correlation with animal protein.

[Table 2 about here]

Table 3 quantifies the discontinuity in the probability of leaving school at age 16 from Figure 1, measuring the difference in the probability of leaving school at age ≥16 for those born before versus in or after 1958. This shows that those born in 1958 onwards are 12 percentage points more likely to stay in school until age 16. With a pre-reform mean of 73%, this suggests that the reform led to a 16% increase in staying in school until age 16. The first stage F-statistic of the preferred specification (column (4)), is 53. Recent work by Lee et al. (2020) shows that the usual critical value of 1.96 for a test at five percent significance level requires the first stage F-statistic to be greater than 104.7. For F-statistics below 104.7, the IV confidence intervals are wider. They show that the critical value for an F-statistic of 53.5 is 2.13. In addition to the standard IV results, I therefore also report the adjusted confidence intervals for all IV estimates, defined as $[\hat{\beta}_1^{IV} \pm 2.13 \times SE(\hat{\beta}_1^{IV})]$. I use the 'standard' 95% confidence intervals for all OLS and reduced form estimates.

[Table 3 about here]

Table 4 shows the results that instrument the measure of education using the dummy for being born in or after 1958.[10] Using a standard 95% confidence interval (not shown here) would suggest that additional education *increases* purchases of calories, total sugars, NMES, starch, fats, saturated fats, and animal proteins. Using the 'tF' procedure from Lee et al. (2020), however, only calories, fats and animal proteins remain significantly different from zero at the 5% level. This shows there is a reduction in the quality of the overall diet, captured by a 0.5 unit increase in the Nutrient Profile Score, though the latter is not significantly different from zero.

[10] Table C.1, Figure C.3 and Figure C.4, Appendix C, show the reduced form effects. Graphs showing the mean of the *covariates* by year of birth show no evidence of discontinuities, suggesting that the individuals observed on either side of the cutpoint are similar in observable characteristics (available upon request).

[**Table 4 about here**]

Looking at the magnitude of the effects, the estimates suggest that leaving school at 16 increases total calories by 619 per person per day, compared to those leaving school earlier. This is a large effect. There are at least three potential statistical explanations for the magnitude of this effect. First, as discussed above, if the higher educated are systematically more likely to purchase nutritionally better foods *within* a food group compared to the lower educated, perhaps because of their higher disposable income, this will lead to the parameter of interest being over-estimated, and hence, this should be interpreted as an upper-bound of the 'true' effect. Second, if the reform led to an increased number of marriages *within* cohorts (e.g. due to assortative mating), the analysis may partially capture spouse's level of education, leading to an upward bias. Due to substantial missing data on spouse's education (33% of our estimation sample), I do not include this in the analysis. As expected, this shows smaller estimates with larger standard errors, but they remain substantial. Third, as the estimates represent a LATE, the coefficients can be interpreted as the causal effect of staying in school until age 16 among those would have left at age 15 had the policy not been introduced. Appendix B shows that the findings are robust to a large set of sensitivity analyses (including controlling for a proxy for income, restricting the sample to ages with common support in treated and control groups, reducing the bandwidth, and controlling for spouse's level of education), though the magnitudes of the estimates for sugar and NMES are less robust.

Nevertheless, given the magnitude of the effects, it is important to consider potential mechanisms that may be driving this. There are at least two mechanisms to highlight. First, the additional income that is associated with staying in school may be used to purchase more foods, thereby increasing the total amount of energy and nutrients purchased among those who stayed in school because of the reform. Second, the higher educated may be more or less likely to drink alcohol and smoke. Both of these health behaviours have been found to affect individuals' nutritional choices, and may explain the estimates found above. I will explore each of these in turn. Following that, I will highlight potential behavioural changes that may offset the negative health consequences of increased food intake.

5.2 Potential mechanisms

One of the mechanisms that may explain the increased nutritional purchases is income. It is well known that the UK schooling reform led to higher wages for those affected (see e.g. Harmon and Walker, 1995; Oreopoulos, 2006). The increase in the amount of nutrients purchased may therefore be driven by increased income for these individuals. That is, if food is a normal good, individuals with higher disposable income may eat more, increasing the total amount of energy and nutrients purchased (unless they have already reached their saturation point). I do not control for income in the main analysis for two reasons. First, being exposed to the schooling reform causally affects earnings, so controlling for such 'downstream' variables may lead to biased estimates of the causal effect of interest. Second, the data do not include any measure of income. However, to explore the sensitivity of the estimates, I follow the literature and rerun the

analyses on nutritional choices controlling for total household spending as a proxy for income (as in e.g., Griffith et al., 2018). The results are presented in Table 5, where Panel A presents the estimates that control for a linear specification of total spending, and Panel B controls for a quadratic in total spending. Both show no substantial changes in the IV estimates, with some additional evidence of an increase in purchases of non-milk extrinsic sugars, suggesting that income does not play a major role in explaining the increase in nutritional choices.

[Table 5 about here]

Other potential mechanisms that may be driving some of the effect of education on nutritional choices are alcohol consumption and smoking. Indeed, alcohol is often considered a complement to food, whereas smoking may be a substitute.[11] This may mean that increased alcohol consumption among higher educated may be accompanied by increased food intake. Similarly, as smoking is more common among the lower educated, this may be one reason why they have reduced food purchases. Table 6 shows the IV estimates of effect of education on different measures of smoking and drinking, using the same specification as above.[12] Columns (1) and (2) show the effect of education on a dummy indicating whether the household spent any money on alcohol, and the expenditures (in £) respectively. Column (3) and (4) show the same for tobacco. The findings indicate that although the IV estimates on alcohol spending are positive and relatively large, the adjusted confidence intervals using the 'tF' procedure renders them insignificantly different from zero at the 5% level. They are marginally significant at the 10% level, providing suggestive evidence that staying in school for longer increases household spending on alcohol. More specifically, staying in school until age 16 increases the probability of positive spending on alcohol by 24 percentage points, or £1.63 per week, on average. Since alcohol and food are considered complements, this suggests that the increased alcohol spending among the higher educated may explain some of the increase in nutrient purchases among this group. I find no significant effects for tobacco spending.

[Table 6 about here]

5.3 Potential offsetting effects

Since the literature that looks at the health effects of additional education either finds no changes or improvements in (physical) health, finding that education reduces the nutritional quality of the diet is somewhat counter-intuitive. Indeed, if education negatively affects morbidity, we would expect to find a positive effect of education on dietary quality. Similarly, if education does not affect morbidity, we would expect to find no effects on dietary quality, unless

[11] For example, evidence shows that some individuals smoke at least in part to control their weight (Cawley et al., 2014), and that smoking causally reduces body weight (van Kippersluis and Rietveld, 2018). There is evidence that the higher educated are more likely to drink alcohol, despite alcohol-related deaths being concentrated among lower SES groups (see e.g. Collins, 2016).

[12] The OLS estimates show that education is positively correlated with alcohol spending, but negatively correlated with spending on tobacco. Both are significantly different from zero at conventional levels. These are not shown here, but are available upon request.

any dietary changes are offset in others ways. I here explore the latter, estimating whether additional education affects other behaviours, potentially mitigating any effects of the reduction in nutritional quality.

More specifically, a reduction in dietary quality does not necessarily affect health if it is offset by changes in physical activity. Indeed, individuals may choose to exercise more to allow them to consume more unhealthy foods, and this may differ by education. Or similarly, higher educated individuals may be more aware of the importance of physical activity, but sustaining a higher level of physical activity requires more foods.

I therefore next explore whether education causally affects individuals' physical activity. For this, I use the Active People Survey (APS) from 2012-2014; a cross-sectional survey of adults in England designed to measure individuals' levels of physical activity. The survey reports individuals' demographic characteristics, and asks them about their physical activity over the four weeks prior to interview, including participation in over 350 types of sports. Individuals are asked to indicate the number of days in the last four weeks they engage in each sport, as well as the usual length of time. In addition, they are asked separately about the amount of walking and cycling they did in the last four weeks, which includes anything even if it is only 10 minutes. I derive the total length of time in each activity by taking the product of the number of days and the usual length of time. I then take the sum of all activities and divide this by the number of days in the month to obtain a measure of the total *daily* minutes of physical activity.

As in the LCFS, I only include individuals born between 1934 and 1982 and I do not observe individuals' month of birth. However, the APS does include the year of birth as well as the month of interview, which allows me to construct lower and upper bounds of the year-month of birth for each individual.[13] I then drop those where it is unclear whether the individual is born before or after 1 September 1957. As the APS only include individuals' highest qualification, and not the age at which individuals left school, I create a binary measure of whether the individual has done their O-levels (an exam usually taken in the year the individual turns 16). The first stage results are consistent with the literature that specifies having O-levels as the variable of interest when exploiting the compulsory schooling reform (see e.g. Dickson et al. (2016)).[14] The final estimation sample includes 168,516 individuals.

The IV results are presented in Table 7 with the reduced form results in Table C.2 and Figure C.5. The latter show that being born after the compulsory schooling reform increased physical activity by approximately 3.2 minutes per day. Scaling this up to the IV estimates, shown in Table 7, suggests that having O-levels increases the amount of time spent in physical activity by just over an hour a day (63 minutes), compared to not having O-levels. This is equally split between time spent in sports (25 minutes a day), and time spent walking (30 minutes a day), though the latter is no longer significantly different from zero when using the

[13]For example, a 50-year-old individual who is interviewed in January 2012 was born between January 1962 and February 1963.

[14]The results show that the reform increased the probability of having O-levels by five percentage points relative to the pre-reform mean of 19%, with a first stage F-statistic of 69. Results are available upon request.

'*tF*' procedure. There is no effect on the time spent cycling. Hence, this suggests that more education, in addition to worsening the quality of dietary purchases, also increases the amount of time individuals spend doing physical activity.[15] Appendix B shows that these estimates are robust to a set of sensitivity analyses, including restricting the sample to ages with common support in treated and control groups, reducing the bandwidth, and controlling for spouse's education.

[Table 7 about here]

One important issue with the above analysis is that it only reflects the amount of physical activity from sports, walking and cycling, whilst there may have been further changes in physical activity due to occupational choice. Indeed, the additional education achieved by those born after the reform may have affected the *type* of job individuals select into. The higher educated, on average, have less physically demanding jobs, so the increased time spent walking and doing sports among the higher educated may be to compensate for the reduced physical activity at work. Unfortunately, the APS does not include information on occupational choice or job-related physical activity.

To explore this in more detail, I use the English Longitudinal Study of Ageing (ELSA), a nationally representative dataset focusing on the dynamics of health, social, well-being and economic circumstances of those aged 50+ in England. ELSA reports participants' occupation using the Standard Occupational Classification 2000 (SOC2000) sub-major groups. To investigate whether the those affected by the reform undertake systematically different amounts of physical activity at work, I assign metabolic equivalents of tasks (METs) to each of the SOC2000 two-digit groups. METs measure the metabolic rate of a particular activity relative to the resting metabolic rate (i.e., the rate when resting, lying down, or sitting quietly). For example, a MET of 2 implies that that activity is twice as strenuous as resting. I assign a MET to all SOC2000 codes in the following way. First, I use the METs assigned to each ISCO-08 code[16] from Deyaert et al. (2017), and convert the ISCO-08 codes to SOC2010 using the mapping from the ONS (2016). Second, I map the SOC2010 to SOC2000 and take the average MET across minor groupings to obtain METs for all two-digit groups, giving me an estimated MET for each ELSA participant based on their SOC2000.[17]

As in the APS, I construct lower and upper bounds of the year-month of birth for each individual. Similar to the LCFS, I use a dummy variable indicating whether the individual left school at age $\geq$16 as the variable of interest. The OLS estimates show that staying in school until 16 is associated with a significant reduction in METs of 0.42 units. In other words, the higher educated are less physically active at work compared to the lower educated. Estimating

[15] This finding is consistent with (Buckles et al., 2013) who find that college graduates are more likely to exercise compared to non-college graduates.

[16] The International Standard Classification of Occupations (ISCO-08) classifies jobs into 436 groups.

[17] There is no 1:1 mapping between SOC2000 and SOC2010, with one of the main revisions in SOC2010 being the reclassifications of managerial functions. I have incorporated these based on the redefinitions provided in (Elias and Birch, 2010). Hence, although there is generally good overlap between the sub-major groupings of SOC2000 and SOC2010 (see e.g. Elias and Birch, 2010), there is likely to remain some incorrect classifications leading to measurement error.

the first stage IV specification, I find almost identical results to those in Table 3: those born after the reform are 13 percentage points more likely to stay in school until age 16. The first stage F statistic is 28.7, and the final estimation sample includes 7,946 individuals.[18]

The IV results are presented in Table 8 with the reduced form results in Table C.3 and Figure C.6, Appendix C. Both analyses show no differences in METs between those born before versus after the compulsory schooling reform. Indeed, the IV estimates in Table 8, suggest that those who stay in school until age 16 have slightly higher METs compared to those who leave school before age 16, but with a large standard error, this is not statistically significantly different from zero.

[Table 8 about here]

5.4 Back-of-the-envelope calculation

The above estimates suggest that education increases the number of calories purchased by approximately 620 kcal per person per day, whilst simultaneously increasing the amount of physical activity by about one hour per person per day, driven by an additional half hour of walking and half hour of doing sports, with no significant changes in job-related METs. To get an idea of the extent of potential offsetting behaviours, I would ideally translate the changes in physical activity into estimates for calories out. However, this is not straightforward, since the measures of education differ between the two datasets used for the analyses. Indeed, the dietary composition analyses estimates the effect of leaving school at 16 or older, whereas the analyses on physical activity estimates the effect of obtaining O-levels. To ensure that the exposure of interest is identical across the different models, I next present some back-of-the envelope calculations based on the reduced form models, reconciling the estimates.

Table C.1 and Table C.2 present the reduced form parameters, showing that being born in or after 1958 increases calorie purchases by approximately 73 kcal per day (this is equivalent to e.g. half a packet of crisps, or 1.5 Oreo cookies), with an increase in physical activity of 3.2 minutes per day (or 1.8 minutes walking and 1.4 minutes of sports).[19] Running the reduced form analysis by gender, I find that men, on average, increase their physical activity by 4.8 minutes per day (2.5 minutes walking and 2.4 minutes of sports). For women, the increase is 2.2 minutes per day (1.4 minutes walking and 0.8 minutes of sports). I can use this to create gender-specific estimates of calories *out* that are driven by the increase in physical activity. I do not create gender-specific estimates of calories *in*, since the data on nutrients are not individual-specific (but rather at the household level; I only estimate an increase in individual energy purchase by scaling it with the 'nutrient equivalence scale').

[18]Note that although ELSA is a panel, I only include one observation per individual, since education and occupation are mostly constant within individuals over time, especially at older ages.

[19]To ensure the estimates for calories in and out are based on a similar sample, I drop the year 1957 in both analyses and control for quadratic annual trends in the year of birth. Including the year 1957, identifying those born before and after September 1957, and specifying year-month trends for the analyses on the APS provides very similar results, which are available upon request.

First, assuming that the average man in the data weighs 189 lbs[20] and walks with a speed of 4mph, they are likely to burn an additional 14 kcal in the extra 2.5 minutes spent walking (relative to resting that time instead). Women burn an additional 7 kcal in the 1.4 minutes spent walking. For the additional 2.4 and 0.8 minutes of sports for men and women respectively, the amount of energy burnt depends how strenuous this exercise is. For light physical activity (e.g. playing with children), the average man and woman burn an additional 11 and 3 kcal respectively (again, relative to resting). For moderate physical activity (e.g. baseball, cricket, golf), it is an additional 17 kcal for men and 5 for women, and for vigorous physical activity (e.g. running), it translates to an additional 31 and 9 kcal for men and women respectively. [21]

Combining these estimates suggests that the minimum increase in the amount of energy expended due to the additional time in physical activity (i.e. with light physical activity) is 25 kcal ($= 14 + 11$) for men and 10 kcal ($7 + 3$) for women. The maximum (i.e. with vigorous physical activity) is 30 kcal ($= 14 + 31$) for men and 16 kcal ($7 + 9$) for women. It is important to note that these are very crude estimates for at least two reasons. First, weight is endogenous to both nutrient purchases and physical activity. And second, the estimates of the amount of energy burned in different activities are only an approximation. Having said that, however, they do suggest that there are some partially offsetting effects in terms of calorie intake and expenditure. Note also that the increase in calories *in* is likely to be over-estimated due to higher educated individuals expected to purchase more expensive or higher quality foods within the same food group. The back-of-the-envelope calculation therefore suggests that at least some of the increase in calories *in* is mitigated by an increase in calories *out* through physical activity, leading to little effect on the balance of calories; consistent with the literature that explores the effect of education on obesity (see Galama et al., 2018).

6 Conclusion and discussion

This paper explores the effects of education on dietary choices, as well as potential compensatory responses in terms of physical activity. To account for the endogeneity of education, I exploit the 1972 UK schooling reform that increased the age at which individuals were legally allowed to leave full-time education from 15 to 16 for those born after September 1957, but not for those born before. The results, perhaps surprisingly, show that increased education *worsens* the overall quality of the diet, causing increases in the purchases of calories, fats, and animal protein. These results are robust across a wide set of model specifications, including the use of an alternative data source, controlling for a proxy for income, restricting the sample to ages with common support in treated and control groups, using different bandwidths, and

[20] Or 86kg. This is the average weight among men born in 1957/8 and observed between 2003–2012 in the Health Survey for England (own calculations). The amount of energy burnt is a function of an individual's body weight.

[21] These numbers are taken from statistics provided by the Harvard Medical School. For example, a 185 lbs individual burns 26 kcal when resting for 30 minutes (or 0.9 kcal per minute). The same individual burns 189 kcal in 30 minutes walking (6.3 kcal p/m). Hence, spending an additional 2.5 minutes walking implies burning between 16 kcal *minus* 2 kcal (i.e. what would be burnt if the individual had rested this time instead), so approximately 14 kcal.

controlling for spouse's level of education (these are shown in Appendix B). However, although the estimates for sugar, non-milk extrinsic sugars, starch and saturated fats are all positive, they are not always precisely estimated, and the results for sugar and non-milk extrinsic sugar are less robust across different specifications in terms of the size of the coefficients.

How do these estimates compare to the literature? Are there other interventions that lead to similar-sized changes in nutritional purchases? This is difficult to say, since much of the literature explores interventions that aim to reduce BMI (rather than those that specifically focus on nutritional intake, see e.g. Cawley et al., 2019), examines changes in quantities consumed (e.g. millilitres per day of sugar-sweetened beverages, see Rayman et al., 2018, or kilograms of fruit and vegetables purchased, see Griffith et al., 2016), or focuses on a particular mealtime/shopping occasion, often without data on individuals' consumption or purchases at other times/elsewhere (see e.g. Bollinger et al., 2011; Downs and Loewenstein, 2012). There are, however, some interesting comparisons. For example, Belot et al. (2018) find some evidence of changes in children's (but not adults') nutritional choices (and BMIs) following two randomised interventions that provided healthy meals or snacks to low-income families. In another randomised experiment, Wisdom et al. (2010) show that making healthy choices more convenient in a fast-food chain led to a reduction in calorie purchases, though this is compensated by increased calories on side orders and drinks, and Fichera and von Hinke (2020) and Dubois et al. (2021) find that nutrition labelling led to a small improvement in the dietary composition of the shopping basket.

In none of these studies, however, is the 'treatment effect' of similar magnitude to what I find here. Indeed, the estimates in this paper remain relatively large in comparison, but there are at least three explanations for this. First, a systematic correlation between education and nutritional quality *within* food groups may lead to an overestimate of the effect of interest in the case of assortative mating. Second, the omission of spouse's education may lead to an upward bias in the estimates. And third, the IV results should be interpreted as a local average treatment effect. Hence, the estimates should be interpreted as an upper-bound of the 'true' effect.

Most of the literature that looks at the health effects of additional education either finds improvements in health, or no changes. Finding that education reduces the nutritional quality of the diet is therefore somewhat unexpected and calls for an investigation into the possible mechanisms. I explore multiple potential explanations. First, education may cause a worsening of the diet through an income effect. More specifically, it is well known that the additional schooling due to the reform increased individuals' wages (see e.g. Harmon and Walker (1995); Oreopoulos (2006)). With more disposable income, these individuals may purchase more foods, increasing the total amount of energy and nutrients purchased. I explore this channel by controlling for total household spending as a proxy for income. The estimates from these analyses remain very similar, suggesting that the increased income associated with the educational reform is not driving my results.

Second, I investigate the role of alcohol and smoking in the relationship between education

and nutrition. Indeed, alcohol is often considered a complement to food, whilst smoking may be a substitute. Increased alcohol intake may therefore coincide with increased food purchases. The analysis shows some suggestion of education increasing spending on alcohol, but not tobacco, which may explain some of the increased nutrient purchases among the group who stayed in school until age 16.

Third, although the higher educated may increase their food *purchases*, they may not necessarily increase their food *consumption* to the same extent. For example, there may be more wastage among higher educated households, or individuals may be more likely to purchase foods for others (e.g. presents or entertaining). I explore this directly in Appendix B by using data from the National Diet and Nutrition Survey, the only source of nationally representative UK data on nutritional *intakes*. The analyses, however, show similar patterns of results, but with substantially smaller samples, the findings are not significantly different from zero.

A fourth potential explanation for the worsening of the diet is that consumers may compensate for healthy food choices using unhealthy ones. For example, Trivedi et al. (2016) find that when individuals purchase healthy foods, they may simultaneously buy a more palatable (but less healthy) food. Hence, if the higher educated spend more on healthy foods, such compensatory behaviours may also increase spending on unhealthy foods, including those high in energy and fat.

Finally, I investigate other behavioural effects that may potentially offset the deterioration in the diet. In particular, I explore the effects of education on physical activity. Indeed, individuals may choose to exercise more to allow them to consume more unhealthy foods (i.e., a licensing effect), and this may differ by education. Or similarly, higher educated individuals may be more aware of the importance of physical activity (i.e., a knowledge effect), but a higher level of physical activity requires more food intake. Results confirm a discontinuity in physical activity by birth cohort, with those who got their O-levels because of the compulsory schooling reform being significantly more physically active per day. Although the analysis on physical activity clearly shows the importance of exploring potential offsetting behaviours, I am unable to identify the exact mechanism (e.g. a licensing vs. knowledge effect) that is driving the deterioration of the diet.

A back-of-the-envelope calculation suggests that the increase in calories *in* is only slightly larger than the increase in calories *out*. Together with the specific measurement error in calories in, this suggests there is little effect on the actual calorie balance. This is consistent with the literature that finds no convincing evidence that education causes changes in obesity. Although I am unable to say which comes first (i.e., do the higher educated eat more because they do more exercise, or do they do more exercise because they eat more?), the analysis highlights the importance of separately exploring the effects on calories in versus calories out, providing additional information that is concealed in the analyses that only use measures of obesity. As such, this study sheds further light on the behavioural responses to an increase in the years of schooling.

One important question relates to what the estimates of education are capturing. First, it

is well known that the UK schooling reform led to higher wages for those affected, potentially leading to an increase in nutrients purchased due to higher disposable incomes. However, I show that the estimates are robust to controlling for a proxy for income. Second, the schooling reform may have led to differential selection into job *types* and changes in employment more generally (see e.g., del Bono and Galindo-Rueda, 2004), with accompanying changes in individuals' time constraints. With the higher educated on average having less strenuous jobs, the reform may have caused a reduction in job-related physical activity, potentially offset by additional discretionary exercise outside of work. I explore this by investigating whether those affected by the reform had different occupational METs, but I find no evidence to support this hypothesis.

Nevertheless, the results in this paper highlight the importance of separately estimating the effects of more disaggregated behaviours, especially when these can be seen as potentially offsetting in their impacts on outcomes such as obesity. Indeed, this paper shows that ignoring the separate subcomponents and only considering the final outcome of interest may conceal different compensating effects.

References

Adda, J. and F. Cornaglia (2006). Taxes, cigarette consumption, and smoking intensity. *American Economic Review 96*, 1013–1028.

Akresh, R., D. Halim, and M. Kleemans (2018). Long-term and intergenerational effect of education: Evidence from school construction in indonesia. *NBER Working Paper 25265*.

Albouy, V. and L. Lequien (2009). Does compulsory education lower mortality? *Journal of Health Economics 28*, 155–168.

Almond, D. and J. Currie (2011). *Human capital development before age 5*, pp. 1315–1486. Elsevier.

Anderberg, D., J. Bagger, V. Bhaskar, and T. Wilson (2019). Marriage market equilibrium, qualifications, and ability. *IZA Discussion Paper 12210*.

Anderson, M. and D. Matsa (2011). Are restaurants really supersizing america? *American Economic Journal: Applied Economics 3*, 152–188.

Atella, V. and J. Kopinska (2014). Body weight, eating patterns, and physical activity: The role of education. *Demography 51*, 1225–1249.

Avendano, M., A. de Coulon, and V. Nafilyan (2017). Does more education always improve mental health? evidence from a british compulsory schooling reform. *HEDG WP 17/10*.

Baker, C. (2018). Obesity statistics. *House of Commons Library Briefing Paper No. 3336*.

Banks, J., L. Carvalho, and F. Perez-Arce (2018). Education, decision making and economic rationality. *CESR-Schaeffer Working Paper No. 003*.

Barcellos, S., L. Carvalho, and P. Turley (2017). Distributional effects of education on health. *mimeo*.

Barreca, A., M. Guldi, J. Lindo, and G. Waddell (2011). Saving babies? revisiting the effect of very low birth weight classification. *The Quarterly Journal of Economics 126*, 2117–2123.

Belot, M., N. Berlin, J. James, and V. Skafida (2018). The formation and malleability of dietary habits: A field experiment with low income families. *IZA discussion paper 11317*.

Bollinger, B., P. Leslie, and A. Sorensen (2011). Calorie posting in chain restaurants. *American Economic Journal: Economic Policy 3*, 91–128.

Buckles, K., A. Hagemann, O. Malamud, M. Morrill, and A. Wozniak (2013). The effect of college education on health. *NBER Working Paper No. 19222*.

Capacci, S. and M. Mazzochi (2011). Five-a-day, a price to pay: An evaluation of the uk program impact accounting for market forces. *Journal of Health Economics 30*, 87–98.

Cawley, J., D. Dragone, and S. von Hinke (2014). The demand for cigarettes as derived from the demand for weight loss: A theoretical and empirical investigation. *Health Economics 25*, 8—-23.

Cawley, J., D. Frisvold, A. Hill, and D. Jones (2019). The impact of the philadelphia beverage tax on purchases and consumption by adults and children. *Journal of Health Economics 67*, 102225.

Clark, D. and H. Royer (2013). The effect of education on adult mortality and health: Evidence from Britain. *American Economic Review 103*, 2087–2120.

Collins, S. (2016). Associations between socioeconomic factors and alcohol outcomes. *Alcohol Research 38*, 833—-94.

Cutler, D., E. Glaeser, and J. Shapiro (2003). Why have americans become more obese? *Journal of Economic Perspectives 17*, 93–118.

Davies, N., M. Dickson, G. D. Smith, G. van den Berg, and F. Windmeijer (2018). The causal effects of education in the uk biobank. *Nature Human Behaviour*.

del Bono, E. and F. Galindo-Rueda (2004). Do a few months of compulsory schooling matter?: The education and labour market impact of school leaving rules. *IZA Discussion Paper 1233*.

Department of Health (1991). Dietary reference values - a guide. Technical report, UK Department of Health.

Deyaert, J., T. Harms, D. Weenas, J. Gershuny, and I. Glorieux (2017). Attaching metabolic expenditures to standard occupational classification systems: Perspectives from time-use data. *BMC Public Health*.

Dickson, M., P. Gregg, and H. Robinson (2016). Early, late or never? when does parental education impact child outcomes. *The Economic Journal*.

Downs, J. and G. Loewenstein (2012). Behavioural economics and obesity. In J. Cawley (Ed.), *Oxford Handbook of the Social Science of Obesity*. Oxford University Press.

Dubois, P., P. Albuquerque, O. Allais, C. Bonnet, P. Bertail, P. Combris, S. Lahlou, N. Rigal, B. Ruffieux, and P. Chandon (2021). Effects of front-of-pack labels on the nutritional quality of supermarket food purchases: Evidence from a large-scale randomized controlled trial. *Journal of the Academy of Marketing Science 49*, 119–138.

Elias, P. and M. Birch (2010). Soc2010: revision of the standard occupational classification. *Economic & Labour Market Review 4*, 48–55.

Epstein, L. and R. Wing (1980). Aerobic exercise and weight. *Addictive Behaviours 5*, 371–388.

Fichera, E. and S. von Hinke (2020). The response to nutritional labels: Evidence from a quasi-experiment. *Journal of Health Economics*.

Fletcher, J. (2010). The effects of soft drink taxes on child and adolescent consumption and weight outcomes. *Journal of Public Economics 94*, 967–974.

Fletcher, J. (2015). New evidence of the effects of education on health in the us: Compulsory schooling laws revisited. *Social Science and Medicine 127*, 101–107.

Fuchs, V. (1982). *Time preference and health: An exploratory study*, pp. 99–120. University of Chicago Press.

Galama, T., A. Lleras-Muney, and H. van Kippersluis (2018). The effect of education on health and mortality: A review of experimental and quasi-experimental evidence. *NBER Working Paper 24225*.

Gelman, A. and G. Imbens (2019). Why high-order polynomials shoul dnot be used in regression discontinuity designs. *Journal of Business and Economic Statistics 37*, 447–456.

Geruso, M. and H. Royer (2018). The impact of education on family formation: Quasi-experimental evidence from the uk. *NBER Working Paper No. 24332*.

Griffith, R., R. Lluberas, and M. Lührmann (2016). Gluttony and sloth? calories, labor market activity and the rise of obesity. *Journal of the European Economic Association 14*, 1253–1286.

Griffith, R. and M. O'Connell (2009). The use of scanner data for research into nutrition. *Fiscal Studies 30*, 339–365.

Griffith, R., M. O'Connell, and K. Smith (2016). The importance of product reformulation versus consumer choice in improving diet quality. *Economica*.

Griffith, R., S. von Hinke, and S. Smith (2018). Getting a healthy start? the effectiveness of targeted benefits for improving dietary choices. *Journal of Health Economics*.

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

Harmon, C. and I. Walker (1995). Estimates of the economic return to schooling for the united kingdom. *American Economic Review 85*, 1278–1286.

Hener, T. and T. Wilson (2015). Marriage market consequences of an educational reform. *mimeo, Ifo Institute*.

Imbens, G. and T. Lemieux (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics 142*, 615–635.

Janke, K., D. Johnston, C. Propper, and M. Shields (2018). The causal effect of education on chronic health conditions. *IZA DP No. 11353*.

Johnston, D., G. Lordan, M. Shields, and A. Suziedelyte (2015). Education and health knowledge: Evidence from uk compulsory schooling reform. *Social Science and Medicine 127*, 92–100.

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

Kemptner, D., H. Jurges, and S. Reinhold (2011). Changes in compulsory schooling and the causal effect of education on health: Evidence from germany. *Journal of Health Economics 30*, 340–354.

King, N., P. Caudwell, M. Hopkins, N. Byrne, R. Colley, A. Hills, J. Stubbs, and J. Blundell (2007). Metabolic and behavioral compensatory responses to exercise interventions: barriers to weight loss. *Obesity 15*, 1373–1383.

Lee, D. and T. Lemieux (2010). Regression discontinuity designs in economics. *Journal of Economic Literature 48*, 281–355.

Lee, D., M. Moreira, J. McCrary, and J. Porter (2020). Valid t-ratio inference for iv. *arXiv 2010.05058*.

Li, J. and N. Powdthavee (2015). Does more education lead to better health habits? evidence from the school reforms in australia. *Social Science and Medicine 127*, 83–91.

Lleras-Muney, A. (2005). The relationship between education and adult mortality in the united states. *Review of Economic Studies 72*, 189–221.

Marmot, M. (2005). Social determinants of health inequalities. *Lancet 365*, 1099–1104.

ONS (2012). Information paper: Quality and methodology information. http://www.ons.gov.uk/ons/guide-method/method-quality/quality/quality-information/economy/summary-quality-report-for-the-living-costs-and-food-survey--lcf-.pdf. [Online; accessed 8-November-2018].

ONS (2016). Soc2010 previous version of the standard occupational classification. https://www.ons.gov.uk/methodology/classificationsandstandards/standardoccupationalclassificationsoc/soc2010. [Online; accessed 13-March-2021].

Oreopoulos, P. (2006). Estimating average and local average treatment effects of education when compulsory schooling laws really matter. *American Economic Review 96*, 152–175.

Rayman, A., L. Jomaa, L. Kahale, P. Adair, and C. Pine (2018). Effectiveness of behavioral interventions to reduce the intake of sugar-sweetened beverages in children and adolescents: a systematic review and meta-analysis. *Nutrition Reviews 76*, 88–107.

Rayner, M., P. Scarborough, British Heart Foundation Health Promotion Research Group, and T. Lobstein (2009). The uk ofcom nutrient profile model. https://www.ndph.ox.ac.uk/cpnp/files/about/uk-ofcom-nutrient-profile-model.pdf. [Online; accessed 19-October-2017].

Silles, M. (2009). The causal effect of education on health: Evidence from the united kingdom. *Economics of Education Review 28*, 122–128.

Trivedi, M., K. Sridhar, and A. Kumar (2016). Impact of healthy alternatives on consumer choice: a balancing act. *Journal of Retailing 92*, 65–82.

van Kippersluis, H., O. O'Donnell, and E. van Doorslaer (2009). Long run returns to education: Does schooling lead to an extended old age? *Journal of Human Resources 4*, 1–33.

van Kippersluis, H. and C. Rietveld (2018). Pleiotropy-robust mendelian randomization. *International Journal of Epidemiology*, 1279–1288.

WHO (2002). *Reducing risks, promoting healthy life: World Health Report 2002*. World Health Organization. [Online; accessed 21-August-2018].

Wilson, T. (2014). Compulsory education and teenage motherhood. *mimeo, University of Glasgow*.

Wisdom, J., J. Downs, and G. Loewenstein (2010). Promoting healthy choices: Information versus convenience. *American Economic Journal: Applied Economics 2*, 164–178.

Figures and Tables

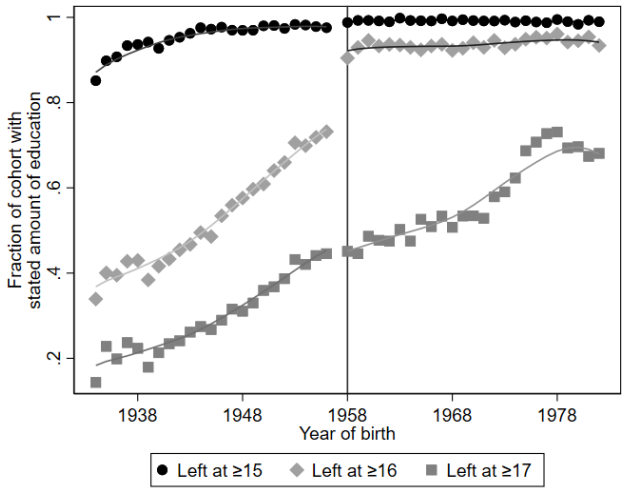


Figure 1: Years of full-time education by birth cohort

Notes: The sample includes all individuals with a year of birth between 1934 and 1982 from the Living Cost and Food Survey, 2003-2015. The vertical line represents the cut-off indicating the first cohort subject to the new compulsory schooling law.

Table 1: Descriptive statistics

	1934-1982	
	mean	sd
Male	0.336	(0.473)
Age	47.992	(12.172)
Married	0.639	(0.480)
Single	0.158	(0.365)
% of total spend (main shopper)	0.829	(0.177)
No. of adults: 2	0.558	(0.497)
No. of adults: 3+	0.152	(0.359)
No. of children: 1	0.148	(0.355)
No. of children: 2+	0.198	(0.399)
North East	0.043	(0.204)
North West	0.115	(0.319)
Yorkshire	0.083	(0.277)
East midlands	0.073	(0.260)
West midlands	0.085	(0.279)
East	0.092	(0.289)
London	0.122	(0.327)
South East	0.134	(0.341)
South West	0.086	(0.281)
Wales	0.049	(0.215)
Scotland	0.089	(0.285)
Northern Ireland	0.028	(0.166)
N	55581	

Notes: Descriptives of the LCFS (2003-2015), including individuals with a year of birth between 1934-1982. All individual-level characteristics refer to the main shopper in the household.

Table 2: OLS results

	(1) Nutient Profile Score	(2) Calories (kcal)	(3) Carbs (g)	(4) Total sugar (g)	(5) NMES (g)	(6) Starch (g)	(7) Fibre (g)	(8) Fats (g)	(9) Saturated fats (g)	(10) Sodium (g)	(11) Proteins (g)	(12) Vegetable proteins (g)	(13) Animal proteins (g)
Left school ≥16	-0.25	-50.76	-9.01	-1.57	-5.24	-13.87	0.68	-3.20	-0.97	-0.13	-0.21	1.12	-1.32
	(0.02)	(16.92)	(1.79)	(1.01)	(0.88)	(2.86)	(0.13)	(0.83)	(0.29)	(0.02)	(0.54)	(0.24)	(0.37)
	[-0.3,-0.2]	[-84.8,-16.7]	[-12.6,-5.4]	[-3.6,0.5]	[-7.0,-3.5]	[-19.6,-8.1]	[0.4,0.9]	[-4.9,-1.5]	[-1.5,-0.4]	[-0.2,-0.1]	[-1.3,0.9]	[0.6,1.6]	[-2.1,-0.6]
Mean	2.0	2609.7	304.5	134.0	89.1	385.0	20.2	103.0	37.8	3.2	84.6	37.9	46.7
Observations	55581	55581	55581	55581	55581	55581	55581	55581	55581	55581	55581	55581	55581

Notes: The table gives the relationship between education and nutritional purchases for a sample of individuals born between 1934 and 1982. All regressions are estimated using pooled waves of the Living Cost and Food Survey 2003-2015. All estimates include a quadratic in the year of birth, a quadratic in age, year and month dummies, gender, marital status, variables indicating the number of adults and the number of children of age 0, age 1, ..., age 17, and region dummies. The confidence intervals are the standard 95% confidence intervals. "Mean" is the mean of the dependent variable (i.e. the average NPS or daily household nutrient purchases) for cohorts born in the 2 years prior to the policy change. NMES stands for non-milk extrinsic sugars. Robust standard errors clustered by cohort in parentheses.

Table 3: First stage IV results

	Pr(Left school ≥16)			
	(1)	(2)	(3)	(4)
I[YOB≥1958]	0.16	0.12	0.12	0.12
	(0.02)	(0.02)	(0.02)	(0.02)
	[0.13,0.19]	[0.09,0.15]	[0.08,0.15]	[0.09,0.15]
1st stage F-stat	95.1	56.0	55.0	53.3
f(YOB)	Yes	Yes	Yes	Yes
D × f(YOB)	No	Yes	Yes	Yes
Age, age^2	No	No	Yes	Yes
Covariates	No	No	No	Yes
Outcome mean	0.73	0.73	0.73	0.73
No. of observations	55581	55581	55581	55581

Notes: The table shows the estimated effect of the 1972 compulsory schooling law change on the probability of leaving school at age ≥16 in the LCFS (2003-2015), including individuals born between 1934 and 1982. The covariates include those mentioned in the note to Table 2. The confidence intervals are the standard 95% confidence intervals. The "Outcome mean" is the mean of the dependent variable for cohorts born in the 2 years prior to the school leaving reform. Robust standard errors clustered by cohort in parentheses.

Table 4: Second stage IV results

	(1) Nutrient Profile Score	(2) Calories (kcal)	(3) Carbs (g)	(4) Total sugar (g)	(5) NMES (g)	(6) Starch (g)	(7) Fibre (g)	(8) Fats (g)	(9) Saturated fats (g)	(10) Sodium (g)	(11) Proteins (g)	(12) Vegetable proteins (g)	(13) Animal proteins (g)
Left school ≥16	0.50	618.93	49.84	34.44	33.48	98.87	-2.91	32.16	10.52	0.48	11.12	-5.91	17.02
	(0.45)	(249.58)	(27.91)	(16.86)	(15.66)	(54.82)	(2.57)	(14.45)	(5.40)	(0.32)	(7.46)	(5.47)	(6.56)
	[-0.5,1.5]	[77.3,1160.5]	[-10.7,110.4]	[-2.1,71.0]	[-0.5,67.5]	[-20.1,217.8]	[-8.5,2.7]	[0.8,63.5]	[-1.2,22.2]	[-0.2,1.2]	[-5.1,27.3]	[-17.8,6.0]	[2.8,31.3]
Mean	2.0	2609.7	304.5	134.0	89.1	385.0	20.2	103.0	37.8	3.2	84.6	37.9	46.7
Observations	55581	55581	55581	55581	55581	55581	55581	55581	55581	55581	55581	55581	55581

Notes: The table gives the IV estimates of education on nutritional purchases for a sample of individuals born between 1934 and 1982, where education is instrumented using the 1972 compulsory schooling law. All regressions are estimated using pooled waves of the Living Cost and Food Survey 2003-2015. The covariates include those mentioned in the note to Table 2. Confidence intervals are calculated using the 'tF' test procedure from Lee et al. (2020). "Mean" is the mean of the dependent variable (i.e. the average daily household nutrient purchases) for cohorts born in the 2 years prior to the policy change. NMES stands for Non-Milk Extrinsic Sugars. Robust standard errors clustered by cohort in parentheses.

Table 5: Controlling for total spending

	(1) Nutrient Profile Score	(2) Calories (kcal)	(3) Carbs (g)	(4) Total sugar (g)	(5) NMES (g)	(6) Starch (g)	(7) Fibre (g)	(8) Fats (g)	(9) Saturated fats (g)	(10) Sodium (g)	(11) Proteins (g)	(12) Vegetable proteins (g)	(13) Animal proteins (g)
Panel A: Linear													
Left school ≥16	0.47	642.71	51.46	35.34	33.77	100.24	-2.63	33.04	10.91	0.51	12.27	-5.10	17.37
	(0.43)	(251.00)	(27.70)	(16.87)	(15.56)	(54.04)	(2.50)	(14.47)	(5.41)	(0.32)	(7.58)	(5.32)	(6.58)
	[-0.5,1.4]	[98.0,1187.4]	[-8.6,111.6]	[-1.3,71.9]	[-0.0,67.5]	[-17.0,217.5]	[-8.1,2.8]	[1.6,64.4]	[-0.8,22.7]	[-0.2,1.2]	[-4.2,28.7]	[-16.7,6.5]	[3.1,31.7]
Panel B: Quadratic													
Left school ≥16	0.55	574.58	45.66	32.62	32.79	93.64	-3.45	30.83	9.97	0.42	9.08	-7.28	16.36
	(0.44)	(262.28)	(28.30)	(17.87)	(16.28)	(54.66)	(2.44)	(15.18)	(5.65)	(0.34)	(8.02)	(5.06)	(7.03)
	[-0.4,1.5]	[5.4,1143.7]	[-15.8,107.1]	[-6.2,71.4]	[-2.5,68.1]	[-25.0,212.2]	[-8.7,1.8]	[-2.1,63.8]	[-2.3,22.2]	[-0.3,1.1]	[-8.3,26.5]	[-18.3,3.7]	[1.1,31.6]
Mean	2.0	2609.7	304.5	134.0	89.1	385.0	20.2	103.0	37.8	3.2	84.6	37.9	46.7
Observations	55581	55581	55581	55581	55581	55581	55581	55581	55581	55581	55581	55581	55581

Notes: See notes to Table 4. Confidence intervals are calculated using the 'tF' test procedure from Lee et al. (2020). Panel A additionally controls for a linear specification of total spending (as a proxy for income); Panel B includes a quadratic in total spending. NMES stands for Non-Milk Extrinsic Sugars.

Table 6: Second stage IV results, alcohol and tobacco spending

	(1) Any spending on alcohol (0/1)	(2) Alcohol spending (£)	(3) Any spending on tobacco (0/1)	(4) Tobacco spending (£)
Left school ≥16	0.24	1.63	-0.06	0.45
	(0.12)	(0.80)	(0.09)	(0.40)
	[-0.0,0.5]	[-0.1,3.4]	[-0.3,0.1]	[-0.4,1.3]
Mean	0.7	2.4	0.3	0.7
Observations	55581	55581	55581	55581

Notes: The table gives the IV estimates of education on purchases of / spending on alcohol and tobacco for a sample of individuals born between 1934 and 1982, where education is instrumented using the 1972 compulsory schooling law. All regressions are estimated using pooled waves of the Living Cost and Food Survey 2003-2015. The covariates include those mentioned in the note to Table 2. Confidence intervals are calculated using the 'tF' test procedure from Lee et al. (2020). "Mean" is the mean of the dependent variable for cohorts born in the 2 years prior to the policy change. Robust standard errors clustered by cohort in parentheses.

Table 7: Second stage IV results, Active People Survey

	(1) Total daily physical activity (in minutes)	(2) Total daily PA (less walking) (in minutes)	(3) Total daily walking (in minutes)	(4) Total daily cycling (in minutes)
O-levels	63.22	24.70	29.65	1.14
	(19.74)	(7.04)	(15.89)	(2.74)
	[22.7,103.8]	[10.2,39.2]	[-3.0,62.3]	[-4.5,6.8]
Mean	45.7	10.3	33.1	2.2
Observations	168516	168516	168516	168516

Notes: The table gives the IV estimates for the effect of education on physical activity (in minutes) for a sample of individuals born between 1934 and 1982. All regressions are estimated using pooled waves of the Active People Survey 2012-2014. All estimates include a quadratic in the date of birth and that interacted with the treatment dummy, a quadratic in age, year and month dummies, gender, and variables indicating the number of adults and children in the household. Confidence intervals are calculated using the 'tF' test procedure from Lee et al. (2020). "Mean" is the mean of the dependent variable for cohorts born in the 2 years prior to the policy change. Robust standard errors clustered by cohort in parentheses.

Table 8: Second stage IV results, English Longitudinal Study of Ageing

	(1) MET
Left school at age ≥16	0.07
	(0.34)
	[-0.7,0.9]
Mean	2.2
Observations	7946

Notes: The table gives the IV estimates for the effect of education on metbolic equivalents of tasks (METs) for a sample of individuals born between 1934 and 1982. All regressions are estimated using one observation per person from the English Longitudinal Study of Ageing. All estimates include a quadratic in the year of birth and that interacted with the treatment dummy, a quadratic in age, year and month dummies, gender, marital status and region dummies. Confidence intervals are calculated using the 'tF' test procedure from Lee et al. (2020). "Mean" is the mean of the dependent variable for cohorts born in the 2 years prior to the policy change. Robust standard errors clustered by cohort in parentheses.

Education, dietary intakes and exercise

— Online Appendix —

Stephanie von Hinke*

* School of Economics, University of Bristol, Priory Road Complex, Priory Road, Bristol, BS8 1TU; Erasmus School of Economics, Erasmus University Rotterdam; and Institute for Fiscal Studies, S.vonHinke@bristol.ac.uk.

Appendix A: Nutritional Profile Model

To measure the nutritional quality, I use the Nutritional Profiling Model that was developed by the UK FSA (Rayner et al. (2009)). This model scores both 'positive' (e.g. fibre) and 'negative' (e.g. sugar) nutrients, and incorporates these into a score indicating the 'healthiness' of the diet.

The nutritional quality is calculated by summing up the number of 'negative' nutrient points, allocated based of the nutritional content per 100g of the shopping basket, and then subtracting the number of 'positive' nutrient points, allocated in a similar way, to obtain a total score. Points are allocated for each of the following individual nutrients: energy, saturated fats, total sugars, sodium, % fruit, vegetables and nuts, fibre and protein. The tables below shows the points for each nutrient, depending on the content per 100g.

Table A.1: Points for 'positive' and 'negative' nutrients

	Points for 'negative' nutrients				Points for 'positive' nutrients		
Points	Energy (kJ)	Saturated fats (g)	Total sugar (g)	Sodium (mg)	Fruit, veg & nuts (%)	Fibre (g)	Protein (g)
0	≤335	≤1	≤4.5	≤90	≤40	≤0.9	≤1.6
1	>335	>1	>4.5	>90	>40	>0.9	>1.6
2	>670	>2	>9	>180	>60	>1.9	>3.2
3	>1005	>3	>13.5	>270	-	>2.8	>4.8
4	>1340	>4	>18	>360	-	>3.7	>6.4
5	>1675	>5	>22.5	>450	>80	>4.7	>8.0
6	>2010	>6	>27	>540			
7	>2345	>7	>31	>630			
8	>2680	>8	>36	>720			
9	>3015	>9	>40	>810			
10	>3350	>10	>45	>900			

To calculate the total score, the model subtracts the 'positive' points from the 'negative' points if the household scores less than 11 'negative' points in total. If the household scores 11 or more 'negative' points, the total score is computed by subtracting the fibre points from the 'negative' score. The FSA defines a food product as 'less healthy' if it scores 4 points or more; a drink is defined as 'less healthy' if it scores 1 point or more.

The original FSA model also subtracts the 'positive' from the 'negative' score if the total 'negative' score is 11 or more with a score of 5 for fruit, vegetables and nuts. The LCSF data do not allow me to distinguish between the fruit, vegetable and nut content.

Appendix B: Robustness analyses

This appendix explores the sensitivity of the IV estimates for the analyses on nutritional choices as well as (if data allow) physical activity. I start by investigating whether the negative effect on dietary quality can be replicated in a different dataset. I then examine the robustness of the findings with respect to (i) restricting the sample to ages with common support in treated and control groups, (ii) the bandwidth in the RD analyses, and (iii) the inclusion of the spouse's level of education.I also examine (iv) whether the effect of education is similar for *food in* versus *food out*, and (v) I investigate any heterogeneity, exploring whether the effect of education on dietary choices is similar for men and women, and those with and without children.

B.1 National Diet and Nutrition Survey

To explore whether the reduction in diet quality is consistent across different datasets, I start by exploring the National Diet and Nutrition Survey (NDNS): a cross-sectional survey recording the types and quantities of foods *consumed* (as opposed to *purchased*) by individuals. These data are available for six years: from 2008/09 to 2013/14. This survey, however, is substantially smaller than the LCFS, with approximately 500 individuals (aged 19+) observed each year. The survey reports individuals' demographic characteristics and asks them to keep a four-day diary recording all food and drinks consumed each day. I use the average (per person) daily consumption of energy and nutrients as the outcomes of interest.

I only include individuals born between 1934 and 1982. Similar to the APS, as year-month of birth is not recorded in the data, I use individuals' year of birth and month of interview to construct lower and upper bounds of the year-month of birth for each individual. I drop those where it is unclear whether the individual is born before or after 1 September 1957.

The first stage IV estimates are almost identical to those using the LCFS: being born in or after September 1957 increases the probability of leaving school at age $\geq$16 by 11 percentage points, from a baseline of 73%. However, with a much smaller sample size (approximately 6.5% of that in the LCFS), the first stage F-statistic is 7.2 (the estimates are not shown here, but are available upon request). The second stage IV estimates are presented in Table B.1, confirming the earlier findings that additional education increases the consumption of energy, carbohydrates, sugars, fibre and protein. However, with the small sample size and large standard errors, none of the estimates are significantly different from zero. Nevertheless, this suggests that the general pattern of results observed for *food purchases* in the LCFS replicates to *food consumption* in the NDNS.

B.2 Common support

The main sample for the analyses on nutritional choices includes all birth cohorts from 1934 to 1982 who are observed in the LCFS between 2003 and 2015, therefore covering a relatively large age range. Although the analyses control for a quadratic in age, it may be that some of the estimates reflect age effects, as opposed to those driven by the schooling reform. I therefore

next restrict the sample to only include ages with common support in both 'treated' (i.e. born in/after 1958) and 'control' (born before 1958) groups in all LCFS waves. In other words, in the 2003 wave, the youngest treated cohort is 45 years old (i.e. born in 1958), with older individuals born before the schooling reform and therefore in the control group. Similarly, in the 2015 wave, the oldest control cohort is 58 years old (i.e. born in 1957), with younger individuals born after the schooling reform and therefore in the treated group. Table B.2 presents the estimates from analysis that restricts the sample to those aged 45 to 58 at the time they were observed. This shows similar estimates, albeit slightly larger. The estimates now also show an increase in purchases of carbohydrates, starch, fibre, sodium, and both vegetable and animal proteins.

B.3 Bandwidth

Identification of the effect of education relies on the treatment being as good as randomly assigned near the cut-point. The smaller the bandwidth, the more likely that this holds. However, reducing the bandwidth also reduces the power, making it more difficult to identify the effects of interest. The main model specification includes 24 birth cohorts on either side of the cut-point. To explore the robustness of the results, Table B.3, Table B.4, and Table B.5 present the findings using a bandwidth of 20 years (Panel A), 15 years (Panel B), 10 years (Panel C), and 5 years (Panel D) for nutritional choices, sports-related and occupational physical activity respectively. Table B.3 shows relatively robust estimates for calories, carbohydrates, fats, saturated fats, and proteins. For example, the estimates show that education increases the number of calories purchased by between 682 and 898 kcal. The results that are less robust though, are sugar and NMES, where the estimates for larger bandwidths show that additional education increases sugar purchases, but this turns negative when the bandwidth is specified at five years (though both are insignificant). This also affects the overall Nutritional Profile Score, which is turns negative with bandwidth of 5 or 10 years.

Table B.4 and Table B.5 shows the results for daily physical activity and METs respectively. The magnitude of the estimates is similar to those in Table 7 and Table 8. For physical activity, however, reducing the bandwidth renders the estimates insignificantly different from zero. For example, using a 5-year bandwidth shows that having O-levels increases the amount of time spent in physical activity by just over 50 minutes a day; equally split between time spent in sports (25 minutes) and time spent walking (22 minutes). There is again no effect on the time spent cycling. The estimates for METs are slightly more variable, but are insignificant in all specifications.

I cannot do this analysis for the APS, since individuals are only interviewed between 2012-2014, restricting the age range with common support to 54–57, leading to a very small sample. However, this common support issue becomes less of a problem when restricting the bandwidth, the results of which I present below. The findings show that, although the estimates are no longer significant due to a lack of power, the effect sizes are very similar in magnitude.

B.4 Accounting for spouse's education

The education reform may have led to complex general equilibrium effects. For example, if it caused an increase in marriages *within* cohorts due to assortative mating, the coefficient on the main shopper's education may partially capture the effect of the spouse's education. The main analysis above ignores the latter, since spouse's education is missing for many observations. Indeed, I do not observe spouse's education for 34% of the estimation sample, leading to a large drop in the sample size. However, omitting this may lead to an upward bias in the coefficient estimates. I therefore explore the robustness of my analysis here, including spouse's education in the estimation.

Similar to the main analysis, I use a dummy variable indicating whether the spouse left school at age 16 or later and I instrument this using a dummy for whether the spouse was born in or after 1958, controlling for a quadratic in their year of birth and age. I additionally drop observations where the spouse is born in 1957 to avoid erroneously assigning individuals to being born before or after the threshold. The estimates are presented in Table B.6. As expected, this shows smaller (though still large) estimates. However, with the substantially larger standard errors and the 'tF' correction, none of the estimates are significantly different from zero.

B.5 Eating in or out

Table B.7 investigates whether education matters more for nutritional purchases *in the home* (Panel A) versus *out of the home* (Panel B). Before interpreting the estimates, it is useful to note that the average household daily nutrient purchases vary substantially between purchases in and outside the home. This is shown in the row indicated "mean" in both panels. With an average nutrient profile score of 1.4 for food in compared to 13 for food out, foods eaten outside the home are substantially *less* healthy. Furthermore, just under one tenth of all energy and nutrients that are purchased by households, are for outside consumption (e.g. 222 kcal for food out versus 2387 kcal for food in).

The OLS estimates of the effect of education show that the higher educated eat more healthily inside the home, but their choices are less healthy away from home. After instrumenting for education, however, it shows that additional education leads to an increase in purchases of energy and nutrients, both inside and outside of the home, though they are not always significant using the inflated 'tF' confidence intervals. For example, those who left school at age 16 because of the reform increased their purchases of energy (kcal) by 545 kcal inside the home, and by 73 kcal outside the home, though the latter is not significant. This suggests that additional education worsens dietary choices more generally, rather than being restricted to either food in or out.

B.6 Heterogeneous effects

Table B.8 and Table B.9 explore whether the effect of education differs by subgroups of the population. Panels A and B of Table B.8 present the estimates for female and male shoppers respectively, showing that additional education increases purchases of calories, NMES, starch, fats, saturated fats, sodium and animal proteins for both genders. However, with the reduced sample size and adjusted confidence intervals, most of these are not significantly different from zero. Any significant effects, however, are only for female shoppers. Table B.9 shows that education increases physical activity for both men and women, but the largest effects are found for men. Indeed the effect sizes for men are double those for women, suggesting that the increase in physical activity for cohorts born after the schooling reform are driven by males.

Panels C and D present the results for households with and without children respectively. Table B.8 shows that the worsening of the diet are driven by the latter, with the results for households with children being generally smaller in magnitude (or negative). However, with the reduced sample size and more conservative confidence intervals, they are generally not significantly different from zero. Although the estimates for physical activity in Table B.9 are larger for families with children, they are not statistically significantly different from zero. Instead, the estimates for families without children suggest they increase their physical activity by 52 minutes per day; 19 of which are due to increased exercise and 24 minutes due to increased walking.

Table B.1: Second stage IV results, National Diet and Nutrition Survey

	(1) Energy (kcal)	(2) Carbohydrates (g)	(3) Total sugar (g)	(4) NMES (g)	(5) Starch (g)	(6) Fibre (g)	(7) Fats (g)	(8) Sats (g)	(9) Proteins (g)
Left school at age ≥16	463.84	100.69	70.31	21.37	30.66	2.15	-0.36	-4.63	22.25
	(434.84)	(72.17)	(44.77)	(31.07)	(39.41)	(5.34)	(19.96)	(9.24)	(22.69)
	[-656.2,1583.9]	[-85.2,286.6]	[-45.0,185.6]	[-58.7,101.4]	[-70.9,132.2]	[-11.6,15.9]	[-51.8,51.0]	[-28.4,19.2]	[-36.2,80.7]
Mean	1712.5	213.6	90.8	50.8	122.7	14.1	69.1	26.1	72.6
Observations	3668	3668	3668	3668	3668	3668	3668	3668	3668

Notes: The table gives the IV estimates of education on nutritional intakes. All regressions are estimated using pooled waves of the National Diet and Nutrition Survey 2008/09-2013/14, and include a quadratic in the year of birth and that interacted with the treatment dummy, a quadratic polynomial in age, year and month dummies, gender, marital status, household size, and regional dummies. Confidence intervals are calculated using the 'tF' test procedure from Lee et al. (2020). NMES stands for Non-Milk Extrinsic Sugars. Robust standard errors clustered by birth year-month in parentheses.

Table B.2: Robustness analyses: Common support, Living Cost and Food Survey

	(1) Nutrient Profile Score	(2) Calories (kcal)	(3) Carbs (g)	(4) Total sugar (g)	(5) NMES (g)	(6) Starch (g)	(7) Fibre (g)	(8) Fats (g)	(9) Saturated fats (g)	(10) Sodium (g)	(11) Proteins (g)	(12) Vegetable proteins (g)	(13) Animal proteins (g)
Left school $\geq$16	-0.71	813.86	84.80	24.58	15.79	214.41	5.52	29.16	9.50	0.94	33.19	14.29	18.90
	(0.58)	(272.05)	(26.27)	(19.54)	(16.39)	(39.88)	(2.47)	(14.78)	(4.99)	(0.37)	(9.84)	(4.15)	(7.60)
	[-2.0,0.5]	[223.5,1404.2]	[27.8,141.8]	[-17.8,67.0]	[-19.8,51.4]	[127.9,300.9]	[0.2,10.9]	[-2.9,61.2]	[-1.3,20.3]	[0.1,1.7]	[11.8,54.5]	[5.3,23.3]	[2.4,35.4]
Mean	2.0	2615.6	305.2	133.9	89.1	389.0	20.3	103.2	37.8	3.2	84.7	38.1	46.7
Observations	17814	17814	17814	17814	17814	17814	17814	17814	17814	17814	17814	17814	17814

Notes: Restricting the sample to ages 45–58. Confidence intervals are calculated using the 'tF' test procedure from Lee et al. (2020). See notes to Table 4. Confidence intervals are calculated using the 'tF' test procedure from Lee et al. (2020). NMES stands for Non-Milk Extrinsic Sugars.

Table B.3: Robustness analyses: Varying the bandwidth, Living Cost and Food Survey

	(1) Nutrient Profile Score	(2) Calories (kcal)	(3) Carbs (g)	(4) Total sugar (g)	(5) NMES (g)	(6) Starch (g)	(7) Fibre (g)	(8) Fats (g)	(9) Saturated fats (g)	(10) Sodium (g)	(11) Proteins (g)	(12) Vegetable proteins (g)	(13) Animal proteins (g)
Panel A: 20y													
Left school ≥16	0.37	785.42	77.10	43.26	38.25	150.01	-0.47	37.15	12.00	0.64	15.12	0.34	14.78
	(0.47)	(291.04)	(30.53)	(21.58)	(19.12)	(51.12)	(2.34)	(16.47)	(6.17)	(0.34)	(8.39)	(4.28)	(7.59)
	[-0.7,1.4]	[153.8,1417.0]	[10.8,143.3]	[-3.6,90.1]	[-3.2,79.7]	[39.1,261.0]	[-5.5,4.6]	[1.4,72.9]	[-1.4,25.4]	[-0.1,1.4]	[-3.1,33.3]	[-8.9,9.6]	[-1.7,31.3]
No. of observations	49697	49697	49697	49697	49697	49697	49697	49697	49697	49697	49697	49697	49697
Panel B: 15y													
Left school ≥16	0.02	682.12	66.32	19.35	18.83	200.33	1.06	35.72	10.03	0.39	13.14	3.84	9.31
	(0.46)	(299.19)	(29.95)	(17.07)	(15.52)	(45.95)	(2.40)	(17.54)	(6.08)	(0.32)	(8.57)	(3.93)	(6.72)
	[-1.0,1.0]	[32.8,1331.4]	[1.3,131.3]	[-17.7,56.4]	[-14.8,52.5]	[100.6,300.0]	[-4.1,6.3]	[-2.3,73.8]	[-3.2,23.2]	[-0.3,1.1]	[-5.5,31.7]	[-4.7,12.4]	[-5.3,23.9]
No. of observations	41301	41301	41301	41301	41301	41301	41301	41301	41301	41301	41301	41301	41301
Panel C: 10y													
Left school ≥16	-0.18	898.41	93.92	24.61	25.55	235.61	3.39	34.90	15.05	1.14	35.91	11.96	23.95
	(0.53)	(255.04)	(25.66)	(17.16)	(18.12)	(43.81)	(2.40)	(13.95)	(5.54)	(0.33)	(10.51)	(3.75)	(9.43)
	[-1.3,1.0]	[345.0,1451.9]	[38.2,149.6]	[-12.6,61.9]	[-13.8,64.9]	[140.5,330.7]	[-1.8,8.6]	[4.6,65.2]	[3.0,27.1]	[0.4,1.9]	[13.1,58.7]	[3.8,20.1]	[3.5,44.4]
No. of observations	28611	28611	28611	28611	28611	28611	28611	28611	28611	28611	28611	28611	28611
Panel D: 5y													
Left school ≥16	-0.50	830.42	117.49	-13.53	-27.06	237.38	7.86	30.03	19.71	0.23	16.24	15.86	0.38
	(0.87)	(621.08)	(90.51)	(37.57)	(31.85)	(163.06)	(4.94)	(34.78)	(15.75)	(0.76)	(15.25)	(11.63)	(10.83)
	[-2.4,1.4]	[-517.4,2178.2]	[-78.9,313.9]	[-95.1,68.0]	[-96.2,42.1]	[-116.5,591.2]	[-2.9,18.6]	[-45.5,105.5]	[-14.5,53.9]	[-1.4,1.9]	[-16.9,49.3]	[-9.4,41.1]	[-23.1,23.9]
No. of observations	13617	13617	13617	13617	13617	13617	13617	13617	13617	13617	13617	13617	13617

Notes: See notes to Table 4. Confidence intervals are calculated using the 'tF' test procedure from Lee et al. (2020). NMES stands for Non-Milk Extrinsic Sugars.

Table B.4: Robustness analyses: Varying the bandwidth, Active People Survey

	(1) Total daily physical activity (in minutes)	(2) Total daily PA (less walking) (in minutes)	(3) Total daily walking (in minutes)	(4) Total daily cycling (in minutes)
Panel A: 20y bandwidth				
O-levels	52.84	29.06	11.58	1.67
	(30.87)	(11.31)	(25.21)	(4.41)
	[-10.6,116.2]	[5.8,52.3]	[-40.2,63.4]	[-7.4,10.7]
Observations	149557	149557	149557	149557
Panel B: 15y bandwidth				
O-levels	38.30	16.32	0.32	2.55
	(36.47)	(12.42)	(30.15)	(5.45)
	[-36.6,113.2]	[-9.2,41.8]	[-61.6,62.2]	[-8.6,13.7]
Observations	119286	119286	119286	119286
Panel C: 10y bandwidth				
O-levels	38.25	4.17	25.39	3.66
	(34.05)	(10.76)	(27.73)	(4.82)
	[-31.7,108.2]	[-17.9,26.3]	[-31.6,82.3]	[-6.2,13.6]
Observations	79972	79972	79972	79972
Panel D: 5y bandwidth				
O-levels	52.32	25.06	21.52	0.19
	(80.68)	(34.16)	(60.29)	(11.76)
	[-113.4,218.0]	[-45.1,95.2]	[-102.3,145.3]	[-24.0,24.3]
Observations	39254	39254	39254	39254

Notes: See notes to Table 7. Confidence intervals are calculated using the 'tF' test procedure from Lee et al. (2020).

Table B.5: Robustness analyses: Varying the bandwidth, English Longitudinal Study of Ageing

	(1) MET
Panel A: Bandwidth of 20 years	
Left school at age $\geq$16	-0.12
	(0.33)
	[-0.9,0.7]
Observations	7247
Panel B: Bandwidth of 15 years	
Left school at age $\geq$16	-0.24
	(0.38)
	[-1.1,0.6]
Observations	6125
Panel C: Bandwidth of 10 years	
Left school at age $\geq$16	-0.44
	(0.30)
	[-1.1,0.3]
Observations	4497
Panel D: Bandwidth of 5 years	
Left school at age $\geq$16	0.43
	(0.58)
	[-0.9,1.8]
Observations	2336

Notes: See notes to Table 4. Confidence intervals are calculated using the 'tF' test procedure from Lee et al. (2020).

Table B.6: Robustness analyses: Accounting for spouse's edcuation

	(1) Nutrient Profile Score	(2) Calories (kcal)	(3) Carbs (g)	(4) Total sugar (g)	(5) NMES (g)	(6) Starch (g)	(7) Fibre (g)	(8) Fats (g)	(9) Saturated fats (g)	(10) Sodium (g)	(11) Proteins (g)	(12) Vegetable proteins (g)	(13) Animal proteins (g)
Left school ≥16	0.19	458.35	40.32	32.72	33.55	119.12	-2.48	22.49	7.74	0.20	9.13	-7.90	17.03
	(0.52)	(307.71)	(36.30)	(18.62)	(16.56)	(104.44)	(2.92)	(15.57)	(5.60)	(0.38)	(8.89)	(6.06)	(8.34)
	[-0.9,1.3]	[-209.4,1126.1]	[-38.5,119.1]	[-7.7,73.1]	[-2.4,69.5]	[-107.5,345.8]	[-8.8,3.8]	[-11.3,56.3]	[-4.4,19.9]	[-0.6,1.0]	[-10.2,28.4]	[-21.0,5.2]	[-1.1,35.1]
Mean	2.0	2608.9	304.6	134.2	89.3	380.4	20.1	103.0	37.8	3.2	84.4	37.7	46.7
Observations	36469	36469	36469	36469	36469	36469	36469	36469	36469	36469	36469	36469	36469

Notes: See notes to Table 4. The analysis also controls for spouse's education (a dummy whether they left school at age ≥16), instrumented by a dummy indicating whether they are born in or after 1958, whilst controlling for a quadratic in the spouse's year of birth and age. IV confidence intervals are calculated using the 'tF' test procedure from Lee et al. (2020).

Table B.7: Robustness analyses: Eating in or eating out, Living Cost and Food Survey

	(1) Nutient Profile Score	(2) Calories (kcal)	(3) Carbs (g)	(4) Total sugar (g)	(5) NMES (g)	(6) Starch (g)	(7) Fibre (g)	(8) Fats (g)	(9) Saturated fats (g)	(10) Sodium (g)	(11) Proteins (g)	(12) Vegetable proteins (g)	(13) Animal proteins (g)
Pabel A: Eating in													
OLS: Left school ≥16	-0.13	-89.32	-13.05	-2.49	-5.74	-10.31	0.40	-5.10	-1.54	-0.19	-1.83	-0.51	-1.32
	(0.03)	(15.90)	(1.71)	(0.98)	(0.85)	(1.00)	(0.12)	(0.80)	(0.28)	(0.02)	(0.52)	(0.20)	(0.37)
	[-0.2,-0.1]	[-121.3,-57.3]	[-16.5,-9.6]	[-4.5,-0.5]	[-7.4,-4.0]	[-12.3,-8.3]	[0.2,0.6]	[-6.7,-3.5]	[-2.1,-1.0]	[-0.2,-0.1]	[-2.9,-0.8]	[-0.9,-0.1]	[-2.1,-0.6]
IV: Left school ≥16	0.23	545.82	43.07	31.56	30.74	11.79	-3.06	30.33	9.90	0.46	9.63	-7.39	17.02
	(0.53)	(241.06)	(27.14)	(16.13)	(14.61)	(18.89)	(2.37)	(14.45)	(5.31)	(0.31)	(7.53)	(4.54)	(6.56)
	[-0.9,1.4]	[22.7,1068.9]	[-15.8,102.0]	[-3.4,66.6]	[-1.0,62.4]	[-29.2,52.8]	[-8.2,2.1]	[-1.0,61.7]	[-1.6,21.4]	[-0.2,1.1]	[-6.7,26.0]	[-17.2,2.5]	[2.8,31.3]
Mean	1.4	2387.3	282.8	124.9	81.6	154.2	19.0	94.3	35.1	2.9	77.4	30.7	46.7
Observations	55581	55581	55581	55581	55581	55581	55581	55581	55581	55581	55581	55581	55581
Pabel B: Eating out													
OLS: Left school ≥16	1.26	38.57	4.04	0.93	0.50	3.04	0.28	1.89	0.57	0.06	1.62		
	(0.11)	(2.98)	(0.27)	(0.13)	(0.12)	(0.16)	(0.02)	(0.12)	(0.03)	(0.00)	(0.09)		
	[1.0,1.5]	[32.6,44.6]	[3.5,4.6]	[0.7,1.2]	[0.3,0.7]	[2.7,3.4]	[0.2,0.3]	[1.7,2.1]	[0.5,0.6]	[0.1,0.1]	[1.4,1.8]		
IV: Left school ≥16	7.12	73.11	6.77	2.88	2.74	3.94	0.15	1.84	0.62	0.02	1.48		
	(1.53)	(36.15)	(3.24)	(1.86)	(1.84)	(3.19)	(0.37)	(1.94)	(0.56)	(0.05)	(1.67)		
	[3.8,10.4]	[-5.3,151.6]	[-0.3,13.8]	[-1.2,6.9]	[-1.3,6.7]	[-3.0,10.9]	[-0.6,0.9]	[-2.4,6.0]	[-0.6,1.8]	[-0.1,0.1]	[-2.1,5.1]		
Mean	13.0	222.4	21.7	9.1	7.5	12.3	1.1	8.7	2.7	0.3	7.1		
Observations	55581	55581	55581	55581	55581	55581	55581	55581	55581	55581	55581		

Notes: See notes to Table 4. OLS confidence intervals are calculated with a 'standard' critical value of 1.96. IV confidence intervals are calculated using the 'tF' test procedure from Lee et al. (2020). Data on the purchases of animal proteins are not available for eating out, so the table only reports purchases of vegetable proteins in Panel B.

Table B.8: Robustness analyses: Heterogeneous effects by subgroups, Living Cost and Food Survey

	(1) Nutrient Profile Score	(2) Calories (kcal)	(3) Carbs (g)	(4) Total sugar (g)	(5) NMES (g)	(6) Starch (g)	(7) Fibre (g)	(8) Fats (g)	(9) Saturated fats (g)	(10) Sodium (g)	(11) Proteins (g)	(12) Vegetable proteins (g)	(13) Animal proteins (g)
A: Female													
Left school ≥16	0.44	487.35	66.97	46.31	38.47	102.45	-1.72	15.55	9.16	0.58	11.81	-5.60	17.42
	(0.43)	(253.75)	(27.77)	(19.43)	(17.73)	(60.05)	(2.53)	(13.05)	(5.24)	(0.37)	(8.74)	(6.38)	(5.66)
	[-0.5,1.4]	[-63.3,1038.0]	[6.7,127.2]	[4.1,88.5]	[-0.0,77.0]	[-27.9,232.8]	[-7.2,3.8]	[-12.8,43.9]	[-2.2,20.5]	[-0.2,1.4]	[-7.2,30.8]	[-19.4,8.2]	[5.1,29.7]
No. of observations	38213	38213	38213	38213	38213	38213	38213	38213	38213	38213	38213	38213	38213
B: Male													
Left school ≥16	0.72	843.15	-10.07	-3.46	16.68	76.16	-6.92	71.53	12.29	0.08	4.98	-7.36	12.33
	(0.98)	(575.22)	(60.38)	(22.70)	(21.28)	(78.36)	(5.16)	(40.54)	(10.50)	(0.67)	(14.61)	(10.08)	(13.99)
	[-1.4,2.8]	[-405.1,2091.4]	[-141.1,121.0]	[-52.7,45.8]	[-29.5,62.9]	[-93.9,246.2]	[-18.1,4.3]	[-16.4,159.5]	[-10.5,35.1]	[-1.4,1.5]	[-26.7,36.7]	[-29.2,14.5]	[-18.0,42.7]
No. of observations	17368	17368	17368	17368	17368	17368	17368	17368	17368	17368	17368	17368	17368
C: Children													
Left school ≥16	0.72	78.66	-29.23	-18.63	-2.61	123.51	-10.94	7.70	2.23	-0.06	-9.45	-14.53	5.08
	(1.04)	(520.16)	(62.79)	(30.26)	(28.74)	(200.57)	(9.77)	(26.83)	(8.55)	(0.63)	(23.03)	(16.70)	(12.43)
	[-1.5,3.0]	[-1050.1,1207.5]	[-165.5,107.0]	[-84.3,47.0]	[-65.0,59.7]	[-311.7,558.8]	[-32.1,10.3]	[-50.5,65.9]	[-16.3,20.8]	[-1.4,1.3]	[-59.4,40.5]	[-50.8,21.7]	[-21.9,32.1]
No. of observations	20049	20049	20049	20049	20049	20049	20049	20049	20049	20049	20049	20049	20049
D: No children													
Left school ≥16	0.52	691.22	61.06	37.63	36.28	103.86	-1.78	33.85	10.46	0.57	14.35	-3.28	17.63
	(0.46)	(325.97)	(38.75)	(19.12)	(15.78)	(54.02)	(2.83)	(17.30)	(6.36)	(0.39)	(8.78)	(4.95)	(7.08)
	[-0.5,1.5]	[-16.2,1398.6]	[-23.0,145.1]	[-3.9,79.1]	[2.0,70.5]	[-13.4,221.1]	[-7.9,4.4]	[-3.7,71.4]	[-3.3,24.3]	[-0.3,1.4]	[-4.7,33.4]	[-14.0,7.5]	[2.3,33.0]
No. of observations	35532	35532	35532	35532	35532	35532	35532	35532	35532	35532	35532	35532	35532

Notes: See notes to Table 4. Confidence intervals are calculated using the 'tF' test procedure from Lee et al. (2020).

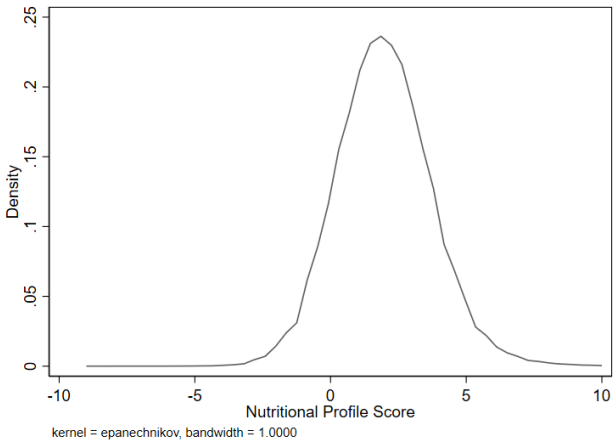
Table B.9: Robustness analyses: Heterogeneous effects by subgroups, Active People Survey

	(1) Total daily physical activity (in minutes)	(2) Total daily PA (less walking) (in minutes)	(3) Total daily walking (in minutes)	(4) Total daily cycling (in minutes)
Panel A: Female				
O-levels	47.40	16.21	26.36	3.07
	(20.47)	(6.61)	(17.44)	(2.37)
	[5.4,89.4]	[2.6,29.8]	[-9.5,62.2]	[-1.8,7.9]
Observations	99698	99698	99698	99698
Panel B: Male				
O-levels	100.89	43.03	37.61	-2.84
	(44.57)	(16.36)	(35.26)	(7.13)
	[9.3,192.4]	[9.4,76.6]	[-34.8,110.0]	[-17.5,11.8]
Observations	68818	68818	68818	68818
Panel C: Children				
O-levels	87.63	27.28	14.85	15.69
	(62.29)	(23.05)	(47.95)	(9.71)
	[-40.3,215.6]	[-20.1,74.6]	[-83.6,113.3]	[-4.3,35.6]
Observations	45453	45453	45453	45453
Panel D: No children				
O-levels	51.81	19.33	24.27	-1.37
	(22.42)	(7.64)	(18.09)	(2.90)
	[5.8,97.9]	[3.6,35.0]	[-12.9,61.4]	[-7.3,4.6]
Observations	123063	123063	123063	123063

Notes: See notes to Table 7. Confidence intervals are calculated using the 'tF' test procedure from Lee et al. (2020).

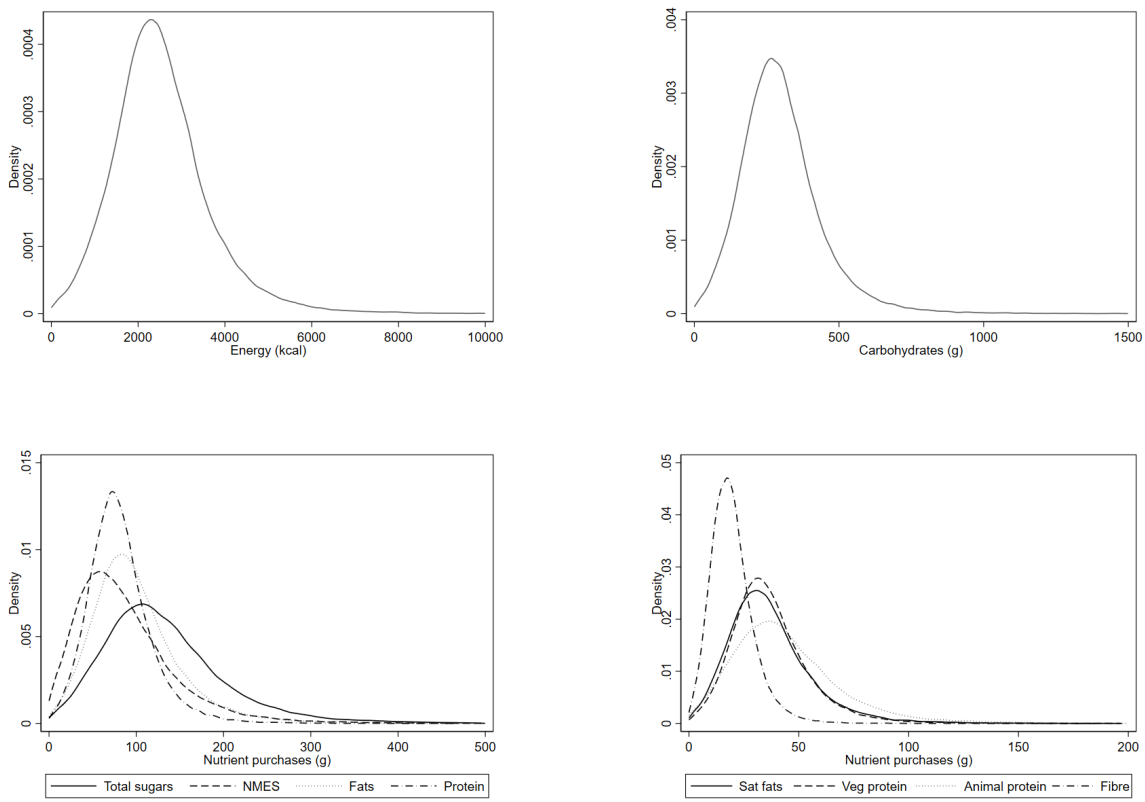
Appendix C: Additional Tables and graphs

Figure C.1: Density of the Nutritional Profile Score



Notes: The figure plots the density of the Nutrient Profile Score from the Living Cost and Food Survey, summarizing the healthiness of households' diets. Each score is based on the total nutrients per 100g of the shopping basket. See Appendix A for more detail.

Figure C.2: Densities of nutrient purchases, 2003-2015



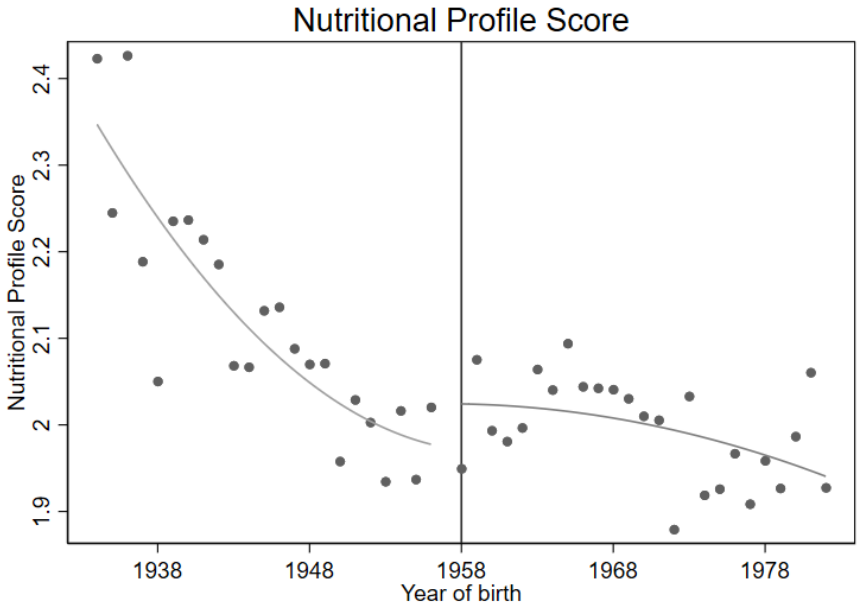
Notes: The figures plot densities of the different nutrients from the Living Cost and Food Survey, where each has been equivalised using the 'nutrient equivalence scale'. Energy is measured in calories (kcal) per person per day, with all other nutrients measured in grams per person per day.

Table C.1: Reduced form results Living Cost and Food Survey

	(1) Nutrient Profile Score	(2) Calories (kcal)	(3) Carbs (g)	(4) Total sugar (g)	(5) NMES (g)	(6) Starch (g)	(7) Fibre (g)	(8) Fats (g)	(9) Saturated fats (g)	(10) Sodium (g)	(11) Proteins (g)	(12) Vegetable proteins (g)	(13) Animal proteins (g)
YOB $\geq$ 1958	0.06	72.89	5.87	4.06	3.94	11.64	-0.34	3.79	1.24	0.06	1.31	-0.70	2.00
	(0.05)	(25.43)	(3.06)	(1.68)	(1.57)	(6.34)	(0.31)	(1.51)	(0.59)	(0.03)	(0.87)	(0.63)	(0.71)
Mean	2.0	2609.7	304.5	134.0	89.1	385.0	20.2	103.0	37.8	3.2	84.6	37.9	46.7
Observations	55581	55581	55581	55581	55581	55581	55581	55581	55581	55581	55581	55581	55581

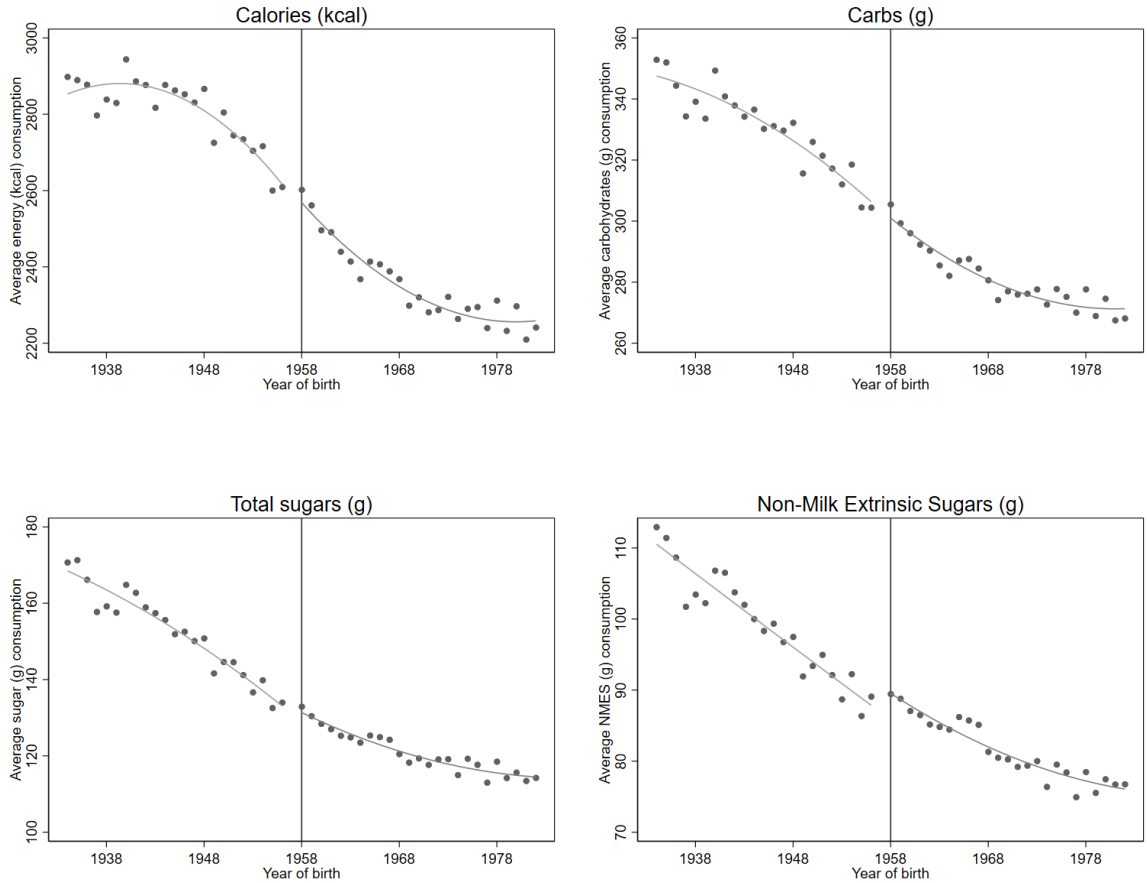
Notes: The table gives the reduced form estimates for the 1972 compulsory schooling law for a sample of individuals born between 1934 and 1982. All regressions are estimated using pooled waves of the Living Cost and Food Survey 2003-2015. All estimates include a quadratic in the year of birth and that interacted with the treatment dummy, a quadratic in age, year and month dummies, gender, marital status, household size, and region dummies. The confidence intervals are the standard 95% confidence intervals. "Mean" is the mean of the dependent variable for cohorts born in the 2 years prior to the policy change. Robust standard errors clustered by cohort in parentheses.

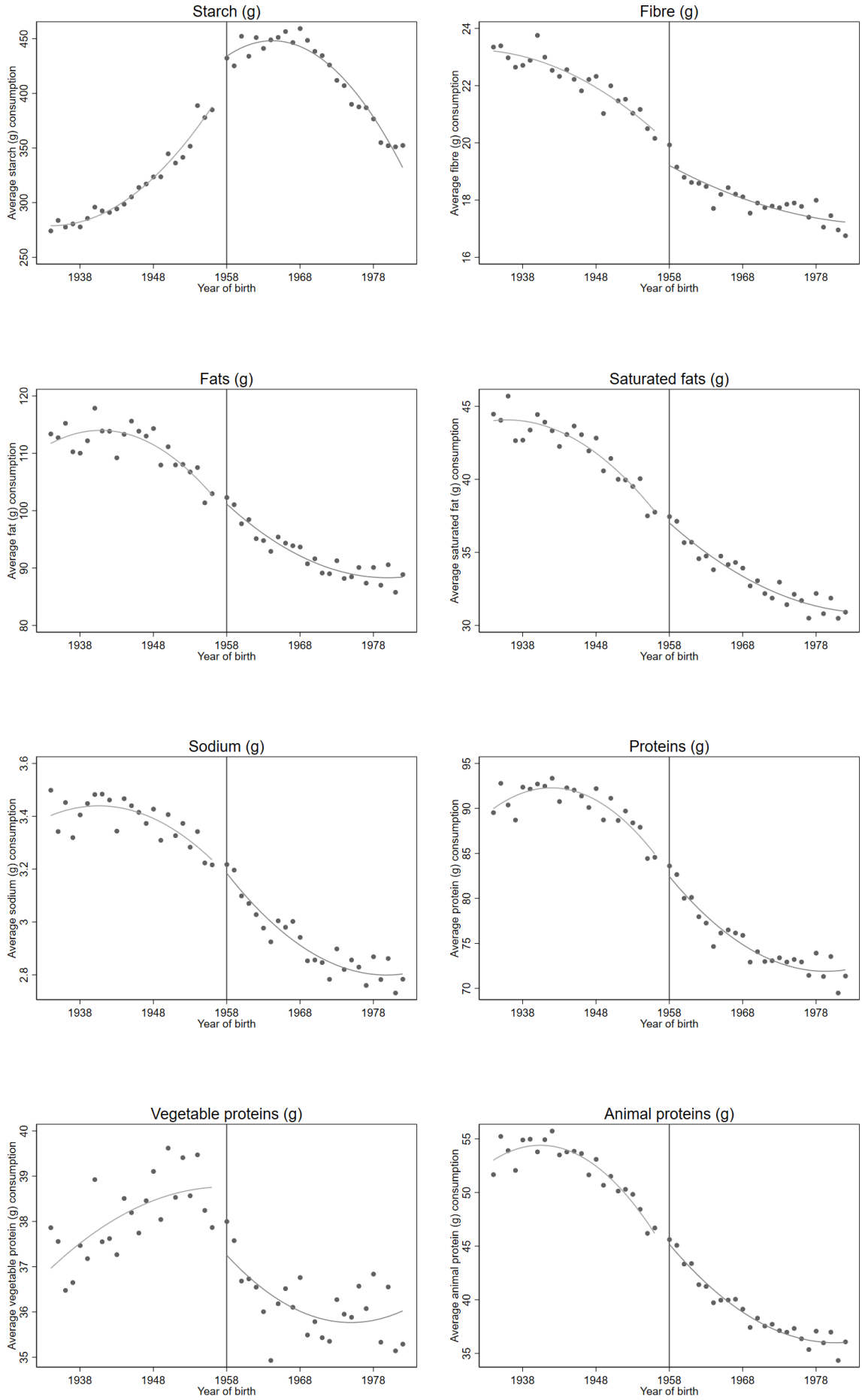
Figure C.3: The Nutritional Profile Score by year of birth, Living Cost and Food Survey



Notes: The figures plot mean values of the Nutient Profile Score on the vertical axis by year of birth on the horizontal axis from the Living Cost and Food Survey, with the vertical line denoting 1958.

Figure C.4: Discontinuities in outcome variables by year of birth, Living Cost and Food Survey





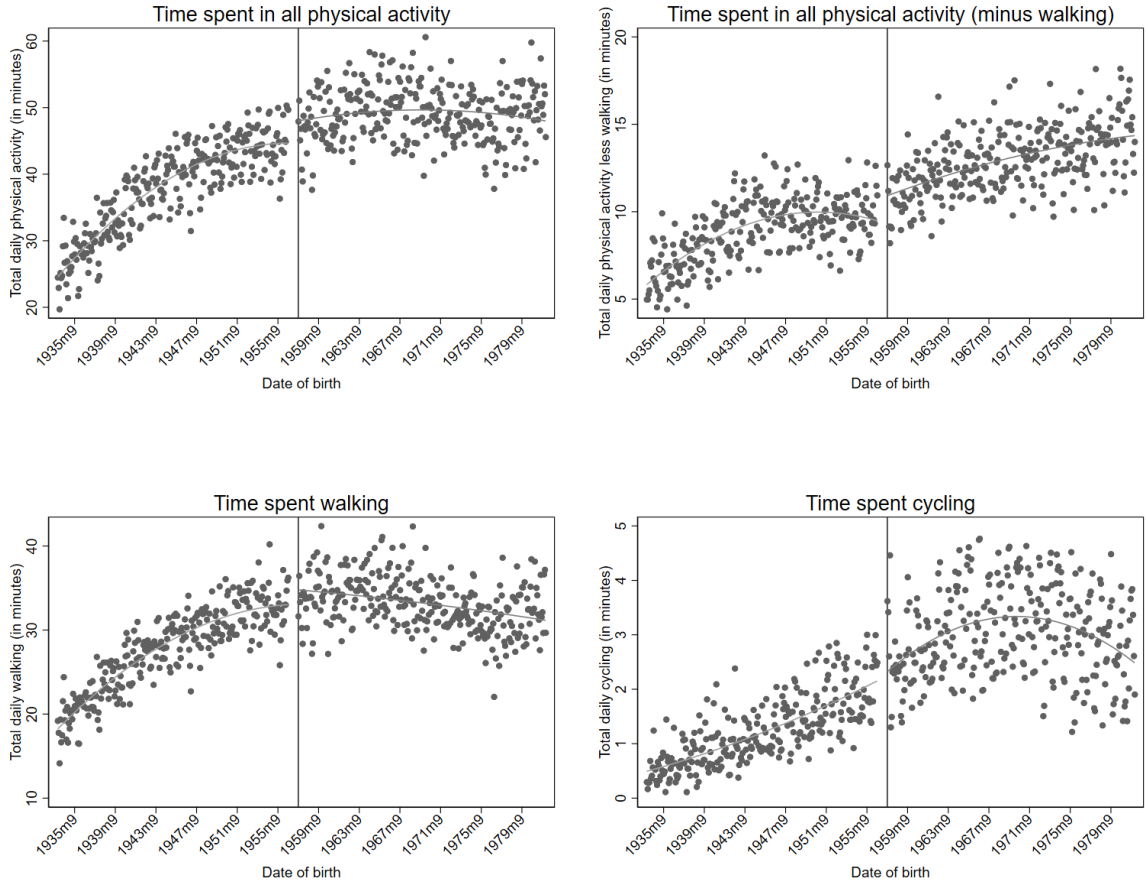
Notes: The figures plot mean values of the dependent variables (measures of nutrition) on the vertical axis by year of birth on the horizontal axis from the Living Cost and Food Survey, with the vertical line denoting 1958.

Table C.2: Reduced form results Active People Survey

	(1) Total daily PA (in minutes)	(2) Total daily PA (less walking) (in minutes)	(3) Total daily walking (in minutes)	(4) Total daily cycling (in minutes)
YOB $\geq$ 1958	3.18	1.44	1.79	-0.06
	(0.82)	(0.39)	(0.62)	(0.12)
	[1.5,4.8]	[0.7,2.2]	[0.5,3.0]	[-0.3,0.2]
Mean	46.8	10.4	33.8	2.6
Observations	328239	328239	328239	328239

Notes: The table gives the reduced form estimates for the 1972 compulsory schooling law for a sample of individuals born between 1934 and 1982. All regressions are estimated using pooled waves of the Active People Survey 2012-2014. All estimates include a quadratic in the year of birth and that interacted with the treatment dummy, a quadratic in age, year and month dummies, gender, and variables indicating the number of adults and children in the household. The confidence intervals are the standard 95% confidence intervals. "Mean" is the mean of the dependent variable for cohorts born in the 2 years prior to the policy change. Robust standard errors clustered by cohort in parentheses.

Figure C.5: Discontinuities in outcome variables by year-month of birth, Active People Survey



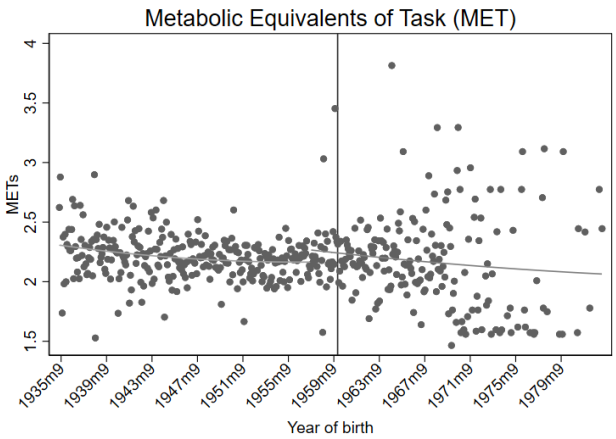
Notes: The figures plot mean values of the dependent variables (measures of physical activity) on the vertical axis by year-month of birth on the horizontal axis from the Active People Survey, with the vertical line denoting September 1957.

Table C.3: Reduced form results English Longitudinal Study of Ageing

	(1) Energy (kcal)
DOB $\geq$ Sept 1957	0.01 (0.04)
Mean	2.4
Observations	7946

Notes: The table gives the reduced form estimates for the 1972 compulsory schooling law for a sample of individuals born between 1934 and 1982. All regressions are estimated using the English Longitudinal Study of Ageing. All estimates include a quadratic in the year of birth and that interacted with the treatment dummy, a quadratic in age, year and month dummies, gender, marital status and region dummies. The confidence intervals are the standard 95% confidence intervals. "Mean" is the mean of the dependent variable for cohorts born in the 2 years prior to the policy change. Robust standard errors clustered by cohort in parentheses.

Figure C.6: Discontinuities in outcome variables by year-month of birth, English Longitudinal Study of Ageing



Notes: The figures plot mean values of the dependent variable (Metabolic Equivalents of Task, or METs) on the vertical axis by year-month of birth on the horizontal axis from the English Longitudinal Study of Ageing, with the vertical line denoting September 1957.