cemmap
centre for microdata methods and practice

Genetic markers as instrumental variables: an application to child fat mass and academic achievement

Stephanie von Hinke Kessler Scholder
George Davey Smith
Debbie A. Lawlor
Carol Propper
Frank Windmeijer

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP03/10

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL
An ESRC Research Centre

Genetic Markers as Instrumental Variables:

An Application to Child Fat Mass and Academic Achievement

Stephanie von Hinke Kessler Scholder [1]
George Davey Smith [2]
Debbie A. Lawlor [2]
Carol Propper [3]
Frank Windmeijer [4]

January 2010

Abstract
The use of genetic markers as instrumental variables (IV) is receiving increasing attention from economists. This paper examines the conditions that need to be met for genetic variants to be used as instruments. We combine the IV literature with that from genetic epidemiology, with an application to child adiposity (fat mass, determined by a dual-energy X-ray absorptiometry (DXA) scan) and academic performance. OLS results indicate that leaner children perform slightly better in school tests compared to their more adipose counterparts, but the IV findings show no evidence that fat mass affects academic outcomes.

Key words: Instrumental Variables; Mendelian Randomization; Genetic Variant; Potential Outcomes; Academic Performance; Educational Attainment; Adiposity; Fat mass; Body Mass Index; ALSPAC
JEL: I1, I2, J24

Acknowledgements:
We thank John Cawley, George Leckie, Owen O'Donnell, Jennifer Roberts, Sarah Smith, Nicholas Timpson, and Gerard van den Berg for helpful suggestions. We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The UK Medical Research Council (MRC), the Wellcome Trust and the University of Bristol provide core support for ALSPAC. This publication is the work of the authors, who will serve as guarantors for the contents of this paper. Funding from three grants supporting the specific work presented here is gratefully acknowledged: two from the UK Economic and Social Research Council (RES-060-23-0011 and PTA-026-27-2335) and one from the UK MRC (G0601625).

[1] *Imperial College Business School, Imperial College London; CMPO, University of Bristol*
[2] *MRC Centre for Causal Analysis in Translational Epidemiology (CAiTE) and Department of Social Medicine, University of Bristol*
[3] *CMPO and Department of Economics, University of Bristol; Imperial College Business School, Imperial College London; CEPR*
[4] *CMPO and Department of Economics, University of Bristol; Centre for Microdata, Methods and Practice, London*

1. Introduction

'Mendelian randomization' refers to studies that exploit the random assignment of an individual's genotype from his/her parental genotypes that occurs at conception (Davey Smith and Ebrahim, 2003). Under certain assumptions that we discuss in detail below, correlations between genetic variants and the outcome of interest cannot be due to reverse causation or confounding by behavioural or environmental factors, including those that occur *in utero*. Mendelian randomization can therefore be exploited to make causal inferences about the effects of modifiable (non-genetic) risk factors, also called phenotypes, on different outcomes (Lawlor et al., 2008).

This approach is receiving increasing attention from economists, several of whom have attempted to identify the causal effect of Body Mass Index (BMI) on economic outcomes (Norton and Han, 2008; Fletcher and Lehrer, 2008; Ding et al., 2009).[1] We contribute to this literature in a number of ways. First, we detail the specific conditions that need to be met for genetic variants to be used as instruments. Genetic epidemiology studies emphasize the importance of carefully examining these conditions,[2] but they have not been well defined in the current economic literature. The increasing availability of biomedical data however, makes understanding of these conditions crucial to the successful use of genotypes as instruments for modifiable risk factors. We therefore discuss the conditions as defined in the epidemiology literature and relate them to the IV assumptions as used in the statistics and economics literature. Second, our empirical application investigates whether child adiposity (fat mass) causally affects academic achievement, using recently identified genetic variants as instrumental variables for adiposity. These variants, in contrast with other genetic variants that have been used in the economics literature, have been shown to be robustly associated with BMI and fat mass in large population samples. Third, recent literature (Burkhauser and Cawley, 2008) has identified the need for a more accurate measure of body size than the commonly used BMI. We use a direct measure of the child's body fat mass, determined by a dual-energy X-ray absorptiometry (DXA) scan. Fourth, our outcome of interest is an objective, independent and comparable assessment of children's achievement. We use the child's score on the UK's nationally set exam taken by all 14-year olds in the English state school system (known as Key Stage 3 tests). Our data further contain rich information on a wide range of family background variables from a large cohort of UK children, allowing us to

[1] BMI is defined as weight in kg divided by height in metres squared.
[2] See for example Davey Smith and Ebrahim (2003) and Lawlor et al. (2008) and references therein.

control for several measures of mother's health and behaviour that may affect both children's body size and educational outcomes, but which are not generally observed in survey data.

When accounting for the endogeneity of child adiposity by using carefully selected genetic variants, we find no evidence of a causal relationship between adiposity and children's academic performance, although the parameters are imprecisely estimated. In contrast, ordinary least squares (OLS) estimates show an inverse association.

The next section introduces our empirical application. Section three presents the estimation strategy, details the conditions that need to be met for genetic variants to be used as instruments, discusses the existing literature and presents our choice of genetic variants. Section four introduces the data. The results are presented in section five; section six concludes and discusses the implications of our findings.

2. Child Adiposity and Academic Achievement

We focus on the effect of children's fat mass on their academic performance. There are three ways in which the two can be related. First, obesity could cause lower academic performance. Second, poor results may cause obesity. And third, instead of there being a causal relationship, the association may be driven by other unobserved characteristics relating to both weight and academic outcomes.

Various pathways have been suggested through which children's adiposity may affect their educational outcomes. Obese children have been shown to be more likely to be absent from school than non-obese children (Geier et al., 2007), which in turn may affect their educational outcomes. Furthermore, associations between obesity and health during childhood may affect educational achievement. For example, obese children are more likely to have sleep apnoea or other sleep disorders (Redline et al., 1999), which are negatively related to cognitive functioning. Another pathway could be that obese children are treated differently by teachers, parents and peers, affecting their (learning) environment (Schwartz and Puhl, 2003). Obese children may be bullied, lowering their self-esteem and harming their educational development. Lower popularity may also lead to ostracism; if this means that, rather than engaging in social activities, children spend more time on their studies, this could lead to better school outcomes. Similarly, fewer recreational activities can increase children's weight and simultaneously increase the time that can be spent on studying (Kaestner and Grossman, 2008).

The reverse causal relationship implies that poor school outcomes cause differences in adiposity, rather than adiposity causing differences in academic performance. Perhaps some children eat excessively to compensate for doing poorly at school. Or conversely, stress caused by poor achievement may lead to reduced appetites and subsequent weight loss (Sabia, 2007).

In terms of unobserved characteristics relating to both adiposity and academic outcomes, even after controlling for an extensive set of background characteristics, there may still be a host of other unobserved family or child factors that are related to both obesity and outcomes. For example, socio-economic position may shape both diet and attitudes to schooling, affecting both weight and school performance. Unobserved time discount rates may be positively related to child weight (i.e. children may be overweight because they place less value on the future) and – with the same reasoning – negatively related to the child's human capital investment and educational outcomes. Similarly, rather than engaging in physical activity, overweight children may have an (unobserved) preference (e.g. level of discipline) to devote this time to studying which in turn increases their academic achievement.

3. The Use of Genetic Variants as Instruments for Child Adiposity

3.1 The Estimation Strategy

We examine the impact of child adiposity at age 11 on their educational outcomes at age 14. We model adiposity as part of a child education production function:

$$S_i = f\left(A_i, X_i, P_i, U_i\right), \tag{1}$$

where S_i, the academic performance of child *i*, is a function of the child's adiposity A_i, a set of child and family background characteristics X_i and parental health and behaviour P_i. The variable U_i represents the unobserved component, which includes both unobserved child attributes and unobserved parental/family behaviour. We start with a simple linear model:

$$S_i = \alpha + \beta A_i + U_i, \tag{2}$$

with β as the parameter of interest: the average relationship between child adiposity and academic achievement. We then augment equation (2) to include the additional sets of

covariates in X_i and P_i, which allows us to explore how the relationship between adiposity and academic achievement changes when controlling for various observed inputs in the child education production function. The possible endogeneity of adiposity is characterised by the fact that the unobservable confounders U_i determine educational outcomes S_i, but also determine adiposity A_i, leading to biased OLS estimates.

We use IV to deal with this, introducing instrumental variables Z_i that are associated with A_i, but only associated with S_i indirectly through its association with A_i. Hence, we estimate (2) and use children's genetic variants as instruments for their fat mass. In the absence of a constant treatment effect, Angrist, Graddy and Imbens (2000) specify the assumptions needed for the standard linear IV estimator in (2) to identify the average causal response within a potential outcomes framework, see also Angrist and Pischke (2009). We briefly outline these assumptions here.

Let S, A and Z denote random variables representing, respectively, the educational outcome, the adiposity measure and the (for now: binary) genetic variant as instrumental variable. Let $S_i(a,z)$ be the potential outcome for individual i that would be obtained if the adiposity measure, the treatment variable, was set to a and the instrument set to z. Equivalently, let $A_i(z)$ be the potential adiposity for individual i when the instrument is set equal to z. We make the following assumptions:

Assumption 1. (Independence and Exclusion)

$$Z_i \perp \{S_i(a,z), A_i(z)\}_{a,z}$$

$$S_i(a,1) = S_i(a,0), \text{ for all } a.$$

Independence means that the instrument is independent of the potential educational outcome and the potential adiposity, for all values of a and z. In other words, the instrument is as good as randomly assigned. Exclusion implies that the potential outcomes, at any level of adiposity a, are unchanged by the presence or absence of the genetic variant.

Assumption 2. (Nonzero effect of instrument on adiposity)

$$E\left[A_i(1) - A_i(0)\right] \neq 0.$$

This implies that expected potential adiposity is affected by the genetic variant and that therefore the coefficient in the (first stage) regression of A_i on Z_i is non-zero.

Assumption 3. (Monotonicity)

$$P\left[A_i(1) \geq A_i(0)\right] = 1,$$

(or vice versa), saying that the potential adiposity for individual *i* with the genetic variant is at least as high as the potential adiposity for the same individual without the genetic variant.

From the exclusion restriction, it follows that $S_i(a,z) = S_i(a)$. Specifying heterogeneous responses, the potential outcome for individual i can be written as a general function of a, say $S_i(a) \equiv g_i(a)$. The IV (or Wald) estimator in equation (2) is then equal to

$$\begin{aligned}\hat{\beta}_{IV} &= \frac{E\left[S_i \mid Z_i = 1\right] - E\left[S_i \mid Z_i = 0\right]}{E\left[A_i \mid Z_i = 1\right] - E\left[A_i \mid Z_i = 0\right]} \\ &= \frac{\int E\left[g_i'(q) \mid A_i(0) < q \leq A_i(1)\right] P\left\{A_i(0) < q \leq A_i(1)\right\} dq}{\int P\left\{A_i(0) < q \leq A_i(1)\right\} dq},\end{aligned}$$

where $g_i'(q)$ is the derivative of $g_i(a)$ w.r.t. a evaluated at q. Therefore, the IV estimator is a weighted average of the derivative function. As Angrist, Graddy and Imbens (2000) show, when the causal response function is linear

$$g_i(a) = \alpha_i + \beta_i a,$$

then

$$\hat{\beta}_{IV} = \frac{E\left[\beta_i \left\{A_i(1) - A_i(0)\right\}\right]}{E\left[A_i(1) - A_i(0)\right]}, \qquad (3)$$

i.e. the IV estimate is a weighted average of the random coefficients β_i, with the weights proportional to the adiposity change induced by the genetic variant. In the case of multiple instruments, $\hat{\beta}_{IV}$ is a weighted average of the Wald estimators using the instruments one at the time. With multi-valued rather than binary instruments, $\hat{\beta}_{IV}$ is a weighted average of the average causal derivatives calculated at each value of the instrument. In both cases, the weights are determined by the relative strength of the instrument in the first stage regression of adiposity on the genetic variants (Angrist and Pischke, 2009; Angrist, Graddy and Imbens, 2000).

3.2 Mendelian Randomization

This paper discusses the use of Mendelian randomization from an economic perspective, with the aim of making causal inferences of the effect of a modifiable risk factor on the outcome of interest. Mendelian randomization however, is closely linked to other study designs used in epidemiology and economics. First, it is linked to the treatment effect literature, where the 'treatment' refers to carrying specific genetic variants. Second, Mendelian randomization is closely related to Randomised Controlled Trials, where the allocation of treatment is randomised over all eligible individuals, as there is an equal probability that either parental allele is transmitted to offspring. Whilst this random allocation is at a family trio level, at a population level it has been demonstrated that genetic variants are largely unrelated to the many socioeconomic and behavioural characteristics that are closely linked with each other and that confound conventional observational studies (Bhatti et al., 2005; Davey Smith et al., 2008; Kivimäki et al., 2008; Lawlor et al., 2008). This therefore suggests that genetic variants are independent of behavioural or environmental factors that may affect the outcome of interest, satisfying Assumption 1 above.

We use Mendelian randomization to make causal inferences about the effect of fat mass on educational attainment. We estimate equation (2), using the child's genetic variants as instruments for its fat mass. Although IV methods are widely used in economics, the use of genetic variants in this field is new. When using Mendelian randomization experiments, genetic epidemiology studies emphasize the importance of carefully examining several situations and (biological) processes that may violate the assumptions mentioned in section 3.1 (see e.g. Davey Smith and Ebrahim, 2003; Lawlor et al., 2008). The existing studies in economics, however, have mainly failed to do so. We therefore begin with a discussion of the

conditions that need to be met to obtain causal estimates of the effect of the risk factor on the outcome of interest.[3] We discuss the concepts defined in the epidemiology literature and relate them to the above assumptions as used in the statistics and economics literature. In this discussion, we focus on our research question: the effect of child adiposity (the modifiable risk factor of interest, sometimes referred to as the intermediate phenotype in Mendelian randomization studies – see Appendix A) on academic achievement (the outcome of interest) and examine the issues that arise in this context.[4]

The first condition relates to Assumption 2, the robustness of the association between the genetic instrument and the risk factor of interest. Mendelian randomization can only be used with genetic variants that have been robustly shown to affect this risk factor. This means it relies on *prior* knowledge about the association between genotype and phenotype, as shown in a large number of independent studies. This latter point is especially important, as many initial genetic associations fail to replicate (Colhoun, McKeigue and Davey Smith, 2003). Without a robust and consistent population association, even if a significant *sample* correlation exists, Assumption 2 may be violated. Any correlation may simply be due to factors such as measurement error (in genotype or phenotype) or chance.

However, even if a suitable and robust genetic instrument is available, it may explain little of the variation in observed phenotype. A weak association could result in a biased estimate and has implications for statistical power. If the alleles shift the adiposity distribution by a very small amount, the effect of adiposity on educational attainment is identified only by this small difference in mean adiposity, emphasizing the need for very large sample sizes, especially when the average causal effect of the risk factor on the outcome could be small. This, of course, is not a problem specific to Mendelian randomization, but refers to a more general problem of weak instruments, often encountered in IV studies (e.g. Angrist and Krueger, 1991).

There can in principle be various situations in which Assumption 1 – that of independence and exclusion of the genetic variants – does not hold, making the instruments invalid. A first situation is that 'behaviours' may be affected by the genotype. As children inherit their genes from their parents, in a study of children's outcomes it is important to consider whether parents' behaviours or preferences are affected by their genotype. For example, mothers who carry 'fat' alleles may be discriminated against in the labour market because of their higher

[3]For a brief description of some key concepts in genetic epidemiology see Appendix A.

[4] For a more general discussion of the use of Mendelian randomization, see Lawlor et al. (2008).

weights (Cawley, 2004). If this affects her behaviour or preferences for her child's education, Assumption 1 may be violated.

A second situation relates to the mechanisms through which genetic variants affect fat mass. These are often unknown.[5] For example, if the mechanism involves changes in behaviour, preferences or brain chemistry that in addition to affecting adiposity also directly affect the outcome, Assumption 1 will be violated. If the mechanisms only result in changes to adiposity (i.e. they do not affect the outcome), Assumption 1 will not be violated.

Thirdly, population stratification refers to a situation in which there exists a systematic relationship between the allele frequency and the outcome of interest in different sub-populations. This can lead to an association between the two at the population level without an actual causal relationship. For example, allele frequencies can vary across ethnic groups. Any systematic differences in educational outcomes across these subpopulations that are not due to a genetic make-up may therefore lead to biased estimates of the effect of adiposity by violating Assumption 1. As a result, it is necessary in Mendelian randomization studies to test whether certain population subgroups are more likely to carry the genetic variant than others.

Fourthly, the genetic instrument may be related to other genetic variants that affect the outcome of interest. Mendel's second law states that the inheritance of one trait is independent of the inheritance of another. However, it has been shown that this does not always hold and that some variants are likely to be co-inherited, especially if they are physically close to each other on a chromosome. Depending on the effects of the co-inherited variant, this so-called 'Linkage Disequilibrium' (LD) can bias the estimates. If our instrument Z is in LD with another polymorphic locus that affects only the phenotype A, the IV estimates remain consistent. However, if Z is in LD with a polymorphic locus that directly affects the outcome S, Assumption 1 is violated and the estimate will be biased and inconsistent. Relatedly, there is the situation of pleiotropy, where one genetic variant has

[5] Until recently, researchers mainly used a 'candidate gene approach' to examine associations between genetic variants (for definitions, see Appendix A) and phenotypes. This approach consists of testing a specific hypothesis: based on biological knowledge, researchers examine the association between one particular genetic variant (the candidate genetic variant) and a phenotype. These studies produced many false-positive findings (Colhoun et al., 2003) and were inefficient. Genome wide association studies (GWAS) followed, genotyping 500,000 to over 1,000,000 SNPs in one go and relating all SNPs to the phenotype of interest in a hypothesis-free way. Stringent criteria are used for GWAS *p*-values to take account of this hypothesis-free approach. Studies are either two-stage studies, where one or more GWAS is performed on samples of individuals, after which the small number of SNPs that reach GWAS levels of statistical significance are typed in other independent samples to examine the extent of robustness. Alternatively, studies consist of a number of independent GWAS containing a large total sample size, where only those SNPs that have consistent associations across all studies are interpreted as robust.

multiple functions. The case is similar to that of LD, and will invalidate the IV approach if the pleiotropic effect influences children's educational outcomes *S* directly, but not if it affects only other characteristics that are unrelated to the outcome of interest.

Finally, another biological process that may bias causal estimates in Mendelian randomization studies is canalisation. This refers to a process by which potentially disruptive influences of the risk factor on the outcome are buffered by foetal or post-natal developmental processes. In other words, where there is evidence during development of differences in an important characteristic (e.g. very high levels of a specific protein) that leads to organs or systems that would be affected by that characteristic to develop differently to avoid adverse consequences. Canalisation may therefore alter the association between genotype and outcome, without any change in the relationship between genotype and risk factor.

With the exception of canalisation, whether Assumption 1 is violated by the conditions outlined above can, in principle, be tested using standard statistical tests of over-identification. However, various studies have directly examined the relationship between genetic variants and individual or family characteristics. These studies provide insight into whether genetic variants are likely to be related to background characteristics or preferences. Davey Smith et al. (2008) for example, estimate pairwise correlations between non-genetic variables and genetic variants and compare the number of correlations that are statistically significant at conventional levels with the number expected by chance if all variables were uncorrelated. This sheds light on the degree of confounding that Mendelian randomization studies may be subjected to.[6] Their findings show 45% of the 4,560 pairwise correlations between behavioural, socioeconomic and physiological factors to be significant at the 1% level. In contrast, genetic variants show no greater association with each other, or with the behavioural, socioeconomic and physiological factors than would be expected by chance. Similarly, to investigate potential selection bias, Bhatti et al. (2005) explore differences in polymorphism frequencies by willingness to participate in epidemiologic studies. They examine three studies with different recruitment designs and different participation incentives. Conditional on having provided blood or saliva samples, they investigate whether genotype frequencies differ by the timing of non-response to questionnaires (early, late and never responders). They find no evidence of correlations between genotypes and response

[6] They use a wide range of non-genetic indicators, such as body size, pulse, vitamin levels, type and frequency of the consumption of various foods, weekly hours of exercise, social class, education, housing tenure, smoking, birth weight, number of siblings, nurse estimation of life expectancy, etc.

characteristics.

With random allocation of genetic variants and the fact that individuals do not know their genotypes, we assume that an individual who carries the risk allele is at least as heavy as the same individual, had she not carried the risk allele, thus satisfying the monotonicity Assumption 3. As this relies on knowing each individual's counterfactual however, this remains an assumption. The literature only tells us that, at a group or population level, those who possess the genetic variant are heavier than those who do not. The monotonicity assumption could for example be violated in the presence of gene-environment interactions.[7]

3.3 The Empirical Evidence using Genetic Variants

The existing economics literature includes three studies that exploit genetic variation to identify the effects of BMI on economic outcomes; they reach different conclusions.[8] Ding et al. (2009) examine the effects of several health conditions, one of which is BMI, on adolescent's academic achievement. Their IV results show large and significant negative effects on female's Grade Point Average (GPA), but not for males. GPAs for obese girls are on average 0.8 points lower than those for non-obese girls. They use four genetic variants as instrumental variables: the dopamine transporter (*DAT1*), the dopamine D2 receptor (*DRD2*), tryptophan hydroxylase (*TPH*) and cytochrome P4502B6 (*CYP2B6*). Fletcher and Lehrer (2008) take a similar approach to Ding et al. (2009), but use a different dataset (the Add Health data) to exploit within-family genetic inheritance with slightly different genetic variants. They find no evidence that obesity affects academic achievement. In addition to *DAT1* and *DRD2*, their instruments include the dopamine D4 receptor (*DRD4*), the serotonin transporter (*5HTT*), monoamine oxidase (*MAOA*) and cytochrome P4502A6 (*CYP2A6*). Finally, Norton and Han (2008) examine the effects of BMI on labour market outcomes using *DAT1* and *DRD4* as instrumental variables for BMI and find no evidence of a causal association.

The discussion in Section 3.2 above highlights the importance of the choice of genetic variants in Mendelian randomization experiments. It states that consistent and robust

[7] Gene-environment interactions occur when the effect of the environment on weight differs depending on the individuals' genetic predisposition, or when individuals' genetic predispositions are expressed differently in different environments. We examine the latter in more detail below.

[8] Our focus is on this literature. Studies that examine similar questions but do not use genetic variation include Sabia (2007), Averett and Stifel (2007), and Kaestner and Grossman (2008). They also report mixed results.

associations should have been shown between the genotype and phenotype in a large number of independent studies. The three economic studies cited above do not appear to have taken this approach (Lawlor, Windmeijer and Davey Smith, 2008). Rather than basing their selection of genetic variants on associations that are robustly shown in the literature, their choice of instruments seems rather ad hoc: using either forward stepwise estimation (Ding et al., 2009) or selecting those SNPs that have statistically significant *sample* correlations in the first stage (Fletcher and Lehrer, 2008). In fact, both Ding et al. (2009) and Fletcher and Lehrer (2008) acknowledge that there is weak and inconsistent evidence in the medical literature, based on very small unrepresentative clinical samples, of the association between their genetic variants and health status or behaviours. Norton and Han (2008) base their selection of SNPs on a study by Guo et al. (2006), who argue that there is a negative association between the D4.7/D4.7 genotype of *DRD4* and obesity. This relationship, however, has not been replicated in other independent studies (see for example Hinney et al. (1999), or Fletcher and Lehrer (2008) who find an insignificant but *positive* association).

In addition, these studies are unable to replicate various associations they note are reported in the literature. For example, Ding et al. (2009) find no association between the number of 10-repeat alleles of the *DAT1* gene and obesity, whilst they note the literature reports a positive relationship, and Norton and Han (2008) find a negative correlation. Fletcher and Lehrer (2008) fail to show any correlation between the A1A1 variant of *DRD2* and obesity. But given that the evidence of a robust association for these variants is lacking, this is not surprising (Lawlor, Windmeijer and Davey Smith, 2008). Furthermore, Norton and Han (2008) argue that the effects of the genetic variants differ by gender, while Patsopoulos et al. (2007) note that most claims of gender differences are spurious. Finally, Norton and Han (2008) use several variants as additional controls rather than instruments, as they fail the over-identification tests (*SLC6A4*, *MAOA*, *DRD2* and *CYP2A6*). Fletcher and Lehrer (2008) and Ding et al. (2009) use several of these as their instruments.

3.4 The Genetic Variants used in the Present Study

We use two SNPs that have been consistently shown to relate to BMI and adiposity in children and adults. Using a total of 38,759 individuals aged between 7 and 80 from 13 different cohorts of European ancestry, Frayling et al. (2007) explore the association between *FTO* and BMI, fat mass, the risk of being overweight and the risk of being obese. They find a

positive association between the risk allele (A) and all measures of adiposity for individuals in all cohorts, in all countries, of all ages and of both sexes, with no difference between males and females. They show that *FTO* is specifically associated with fat mass, with weaker associations with lean mass. In addition, there is no association between *FTO* and birth weight, or *FTO* and height, suggesting that the relationship with BMI is largely driven by individuals' adiposity. They find that each copy of the risk allele is associated with an average increase in BMI of 0.2 units at age 7, to 0.4 units at age 11. For the average age-specific height, this refers to a weight increase of 0.3 and 0.9 kg respectively. As the genetic model for *FTO* is additive, meaning that each risk allele affects the phenotype by a similar amount, 11-year-olds who are homozygous for the rare allele (AA) are on average 1.8 kg heavier compared to those carrying no rare alleles. Using age-specific growth charts of weight, this corresponds to an increase from the median to the 58^{th} percentile. However, there is much variation around this mean effect; the R^2 of a (linear) regression of adiposity on *FTO* is less than 1%.

Several different SNPs near *MC4R* have been associated with adiposity. We use the SNP identified by Loos et al. (2008). In addition to replicating the *FTO* findings, they find a positive relationship between rare allele (C) of *MC4R* and adiposity in genome-wide association data from 16,876 individuals and confirm this relationship in an additional 60,352 adults and 5,988 children. They find no differences by gender, and no effects on birth weight or children's height, again suggesting the association is mediated largely through an effect on adiposity. The genetic model for *MC4R* is dominant, meaning that the presence of any risk allele – either one or two – is associated with a similar increase in adiposity (Timpson et al., 2009). The findings on both *FTO* and *MC4R* have since been replicated in other studies and meta-analyses (see e.g. Willer et al., 2009).

Our choice of genetic variants can be related to the assumptions for suitable use of genetic variants as instruments discussed in Section 3.1 and 3.2. First, the prior findings of robust associations between the genetic variants and individual adiposity justify their use as instruments (Assumption 2). Each *FTO* risk allele leads to an average increase of 0.9 kg; carrying one or two *MC4R* risk alleles is related to an average increase of 0.6 kg. As mentioned above however, with much variation around this mean effect, our two genetic variants explain little of the total variation in adiposity: $R^2 < 1\%$. Using the standard statistical tests, we will examine the strength of our instruments in the application below. Second, we are able to test whether our genetic variants are correlated with a wide range of

maternal characteristics and behaviours (Assumption 1). We also examine whether our variants are associated with various child and family background characteristics by testing whether there are specific patterns in observable characteristics between homozygotes for the common allele, heterozygotes and homozygotes for the rare allele. In this exercise, we examine all covariates used in our analysis as well as an additional random set of background variables that are not included in our specifications. The latter will provide further evidence that our instruments are likely to satisfy Assumption 1. As discussed in Section 2 however, even after examining a broad range of background indicators, there may still be other variables unobserved to the researcher. We therefore rely on the theory of random allocation of genetic variants and on the empirical evidence that shows that genetic variants are unlikely to be related to unmeasured confounders (Bhatti et al., 2005; Davey Smith et al., 2008; Kivimäki et al., 2008; Lawlor et al., 2008). Third, the possible mechanisms through which *FTO* and *MC4R* affect adiposity are increasingly studied in the medical literature. Although this work is ongoing, the current evidence suggests that the variants are associated with an increased consumption of fat and energy (Timpson et al., 2008). The literature suggests that the SNPs increase food intake due to diminished satiety (Wardle et al., 2008), rather than through pathways that affect our schooling outcome of interest, suggesting that Assumption 1 is satisfied.[9]

Fourth, as noted in section 3.2, genetic confounding may occur due to population stratification, LD or pleiotropy. Although *FTO*-allele frequencies are known to vary by ethnic group (Frayling et al., 2007), population stratification due to ethnicity is not likely to be a problem in our data, as our cohort is recruited from a specific geographically defined region with a predominantly white population. In our analysis however, we only include children whose mother describes herself and the child's father as white. Pleiotropy or LD would bias the IV estimates if the variant affects the outcome directly or if the linkage is with another variant that directly affects educational attainment. For example, if the genetic variant used as the instrument directly affects IQ or is in LD with a variant that affects IQ, the IV estimate of the causal effect of adiposity on educational outcomes would be biased. LD however, is not likely to occur for genetic variants on different (non-homologous) chromosomes, and the degree of linkage is a function of the distance between the loci. For our genetic variants to be in LD with an 'ability marker', they would therefore have to sit on the same chromosome and

[9] In addition, mice studies have shown inactivation (knock-out) of the gene to be associated with increased energy expenditure, despite a decrease in locomotor activity (Fischer et al., 2009).

be physically close to each other. Similarly, if *FTO* and *MC4R* are markers for – say – sleeping problems or self-esteem and *through* that affect both adiposity and outcomes, they would be invalid instruments. In Appendix B, we examine the relationship between our variants and a variety of factors that may affect educational outcomes. These show no evidence of a consistent association between our variants and a wide range of factors, suggesting that Assumption 1 is satisfied. Finally, as noted above, the mechanisms through which *FTO* and *MC4R* affect adiposity are unknown, but the evidence to date suggests that size at birth is not affected by these variants. Therefore, canalisation during the foetal period is unlikely to be a problem here.

4. Data

Our data are from a cohort of children born in one geographic area (Avon) of England. Women eligible for enrolment in the population-based Avon Longitudinal Study of Parents and Children (ALSPAC) had an expected delivery date between 1 April 1991 and 31 December 1992. Approximately 85% of these mothers enrolled, leading to about 14,000 pregnancies. The Avon area has approximately 1 million inhabitants and is broadly representative of the UK as a whole, although slightly more affluent than the general population (Golding et al., 2001).[10] Detailed information on the children and their families has been collected using a variety of sources, including self-completed questionnaires, data extraction from medical and educational records, in-depth interviews, and biological samples. Hence, the variables in ALSPAC relate to an unusually wide range of child health and development indicators, family background, family inputs and school measures.

A total of 12,620 children survived past the age of 1 and returned at least one questionnaire. Of these, 642 were excluded because either their mother or father is of non-white ethnic origin, leaving 11,978 potential participants. Our sample selection process is as follows. First, we select those children for whom we observe their genotypes, leaving us with 7,368 children. Second, we drop children with missing data on fat mass. Children were invited to attend specially designed clinics, where their anthropometric measures were recoded. As not all children attended these clinics, our sample sizes reduce to just over 4,500. We further restrict the sample to those children for whom we observe their educational outcomes,

[10] See www.bristol.ac.uk/alspac for more a detailed description of the representativeness of the sample, its enrolment, and response rates.

leading to a final sample size of 3,729 children.[11] We deal with missing values on other covariates using multivariate imputation. We also discuss the results after imputing missing values for all variables apart from the genotypes, resulting in a sample size of 7,368 children.

4.1 Measures of Academic Achievement

Our main outcome measure is the child's Key Stage 3 (KS3) score. The KS3 exam is a nationally set exam, taken by all 14-year-olds in English state schools.[12] This measure of children's performance is therefore objective and comparable across all children. Children's scores for three subjects (English, maths and science) are obtained from the National Pupil Database, a census of all pupils in England within the state school system, which is matched into ALSPAC. We use an average score for the three subjects, standardised on the full sample of children for whom data is available, with mean 100 and standard deviation 10.

4.2 Measures of Child Adiposity and the Genetic Variants

Our main measure for child adiposity is the child's body fat mass (adjusted for age in months, height and height squared), as measured by a dual-energy X-ray absorptiometry scan (DXA) at age 11. This method scans the whole body, dividing it into body fat, lean tissue mass, and bone density. We standardise fat mass on the full sample of children for whom data are available, with mean 100 and standard deviation 10. For the genetic variants, we use two SNPs that have been consistently found to relate to weight: *FTO* (rs9939609) and *MC4R* (rs17782313).[13] Due to the nature of the association between *MC4R* and adiposity (a dominant genetic model), we specify this as a binary variable indicating whether the child carries at least one risk allele (C); i.e. we compare individuals with genotype CC or CT to those with genotype TT. *FTO* is specified as having three categories: no risk alleles (homozygous TT), one risk allele (heterozygous AT) and two risk alleles (homozygous AA).

[11] *t*-tests of mean equality show the final sample to be slightly wealthier than the original ALSPAC sample, with mothers being somewhat older and having fewer mental health problems. The probability of being in the final sample however, is unrelated to *FTO* and *MC4R*, suggesting that sample selection is unrelated to the genotypes measured here.

[12] 93 percent of English children attend state schools.

[13] The rs-number (reference SNP, or RefSNP) is an identification tag that uniquely positions the polymorphism in the genome. All genotyping was performed by KBioscience (http://www.kbioscience.co.uk). SNPs were genotyped using the KASPar chemistry, which is a competitive allele-specific PCR SNP genotyping system using FRET quencher cassette oligos (http://www.kbioscience.co.uk/genotyping/genotyping-chemistry.htm).

4.3 Control Variables

We observe an unusually rich set of child and family background characteristics that we include as covariates as they may be related to both adiposity and the child's educational performance. We control for the child's birth weight and for the number of older and younger siblings under 18 in the household. As children's educational outcomes are known to differ with their age, the analyses include binary indicators for children's age (in months). We include several controls for socio-economic position: log equivalised family income and its square, four binary indicators for mother's educational level, the mother's parents' educational level, an indicator for whether the child is raised by a lone parent, binary indicators for the family's social class, maternal age at birth, and parents' employment status when the child is 21 months. We also include a measure of small (local) area deprivation: the Index of Multiple Deprivation (IMD).[14]

In addition to these generally observed controls, our data allow us to also account for further measures of mother's health and behaviour, which may be correlated to both children's adiposity and educational attainment. In addition, we use these to test whether mother's behaviour differs significantly for the different genotypes. We include two binary variables which measure whether the mother smoked or drank alcohol in the first three months of pregnancy and account for ordered indicators for the intensity of mother's breastfeeding (never, <1 month, 1-3 months and 3+ months). We include the mother's 'locus of control', a psychological concept that describes whether individuals attribute successes and failures to internal or external causes. Those with an internal (low) locus of control see themselves as responsible for the outcomes of their actions. Those with an external (high) locus of control believe that successes and failures are chance-determined. Although the literature does not agree about whether maternal mental health affects the child's cognitive functioning (see e.g. Petterson and Albers (2001) and Frank and Meara (2009)), we include two measures to account for possible confounding.[15] Finally, we use several measures of parental involvement

[14] Family income is an average of two observations (when the child is aged 3 and 4) and is in 1995 prices. The educational indicators are: less than ordinary (O) level, O-level only, advanced (A) level that permits higher educational study, and having a university degree. We use the standard UK classification of social class based on occupation (professional (I), managerial and technical (II), non-manual skilled (IIInm), manual skilled (IIIm), semi-skilled (IV) and unskilled (V)). IMD is based on six deprivation domains, including health deprivation and disability; employment; income; education, skills and training; housing; and geographical barriers to services. Increasing IMD scores indicate greater deprivation. The IMD measure relates to areas containing around 8000 persons.

[15] These are the Edinburgh Post-natal Depression Score (EPDS) and Crown-Crisp Experimental Index (CCEI) at 18 weeks gestation. EPDS indicates the extent of post-natal depression; CCEI captures a broader definition of

or interest in the child's development[16], and a continuous indicator measuring the extent to which the parents engage in active (outdoor) activities with their children.[17]

4.4 Descriptive Statistics

We begin by examining whether the instruments are associated with adiposity by plotting the empirical fat mass distribution functions (as in Angrist, Graddy and Imbens, 2000) for individuals who are homozygous for the common allele (TT), heterozygotes (TA) and homozygotes for the rare allele (AA) for *FTO*. Similarly, we plot the distribution functions for those who are homozygous for the common allele (TT) or either heterozygous or homozygous for the rare allele (TC or CC) for the *MC4R* variant. The left panel of Figure 1a shows large differences between the distributions, with homozygotes for the rare allele of *FTO* having most fat mass. The right panel of Figure 1a presents the differences between these distribution functions, plotting the unnormalised weight functions. As discussed in section 3.1, the IV approach estimates the coefficient of interest (3) as the average causal derivative for the shift in adiposity at each value of the instrument. We therefore show the weight functions for those combinations of Z used in the IV estimation: TA vs. TT (solid line) and AA vs. TA (dashed line).

The figure shows very similar weight functions for the two *FTO* genotypes, both affecting an equally large area of the fat mass distribution (between the values 85 and 130). Only the very top and bottom of the distribution is not captured by the instruments due to the small sample sizes (only 70 children have fat mass values below 85 and 40 children have values above 130). Figure 1b shows the graphs using the *MC4R* indicator. With a dominant genetic model, *MC4R* is specified to be binary. The right panel of Figure 1b shows smaller weights for *MC4R* compared to *FTO*, reflecting the smaller change in adiposity induced by *MC4R*.

Appendix C presents summary statistics for individuals who are homozygous for the common allele, heterozygous and homozygous for the rare allele of *FTO* and *MC4R*.[18] The

---

mental health, measuring general anxiety, depression and somaticism. Higher scores mean the mother is more affected.

[16] A continuous variable ranging from 0-10 is included measuring the mother's 'teaching score'. This is constructed from questions that measure whether the mother is involved in teaching her child (depending on the child's age) songs, the alphabet, being polite, etc. We use an average score from three measures at ages 18, 30 and 42 months to capture longer-term involvement. Likewise, a variable is included indicating whether the mother reads/sings to the child, allows the child to build towers/other creations etc, measured at age 24 months.

[17] This includes several recreational pursuits (including going to the park or playground and going swimming).

[18] Rather than using a binary *MC4R* indicator, we distinguish between the three categories to investigate possible associations between the number of risk alleles and all other covariates, although the findings are

first two rows show significant relationships between the number of risk-alleles of either *FTO* or *MC4R* and fat mass or BMI. The table shows that these differences are mainly driven by children's weights rather than their heights, which shows no large differences for both variants.

Finally, the contextual variables and the measures of mother's health and behaviour do not show any clear patterns of differences across the child's genotypes.[19] For example, those who are heterozygous or homozygous for the *FTO* risk allele are of slightly higher social class, and mothers whose child is heterozygous or homozygous for the *MC4R* risk allele are less likely to work full-time. However, there is no obvious structure in the magnitude or significance of these differences, supporting the assumption that the genotypes are distributed randomly in the population and are unrelated to other child or family characteristics. Also note that the two genetic variants are uncorrelated (*r=-0.006*), and that both variants lead to an increase in adiposity. If this rise in adiposity affects (e.g.) mother's behaviour, we would expect to see similar patterns or differences in mean characteristics for both variants. Only for mother's locus of control do we find increasing scores for both variants; none of the other behaviours or preferences show similar patterns, supporting the assumption that they are not affected by the genotype (condition two in section 3.2 above).[20]

5. Results

5.1 OLS

Table 1 presents the OLS results from the regression of the KS3 scores at age 14 on fat mass at age 11.[21] The relationship between fat mass and educational attainment when not controlling for any other covariates – equation (2) above – is negative, with a one standard deviation increase in fat mass associated with a 0.1 standard deviation decrease in test scores (column 1). We augment this to account for the contextual variables (column 2) and mother's

similar when using a binary variable. Robust patterns of significant differences would indicate possible violations of Assumption 1.

[19] There are also no differences in mother's health and behaviour by mother's genotype (results available from the authors).

[20] Appendix B presents the relationship between our SNPs and an additional set of background variables that are not included in our analyses. Apart from strong positive associations with waist and hip circumference, it shows no clear patterns of significant differences, providing further evidence that our instruments are likely to satisfy assumption 1.

[21] A non-parametric locally weighted regression between educational attainment and the child's fat mass indicates a clear negative relationship, linear over the full range of the fat mass distribution.

health and behaviour (column 3).[22] This brings the estimate closer to zero, but a negative association remains. The estimated coefficients of the other covariates (not shown) are in line with priors and other analyses of educational attainment on these UK tests.[23] In summary, the OLS findings suggest that there is an inverse correlation, albeit a small one, between children's fat mass and their educational attainment.

5.2 IV Estimates

We now turn to the IV estimates, instrumenting the child's fat mass with the two genetic variants. As mentioned above, *FTO* contains three distinct values, indicating the number of risk alleles carried by the child, which we include as a linear indicator. *MC4R* is a binary variable indicating whether the child is either heterozygous or homozygous for the rare allele. Table 2 presents the first-stage regression results, showing a strong positive relationship between *FTO*, *MC4R* and child fat mass. As we move across the columns, more controls are added to the model, with no significant changes in the magnitude or sign of the instrument-coefficients. The strength of the relationship between the instruments and fat mass is shown by the first stage *F*-statistic; with values above 20, it is not a weak instrument and passes the standard statistical test. In fact, if we were to use only the stronger *FTO*, the *F*-statistic rises to values over 35, confirming the strength of our genetic variants.[24]

The IV results are presented in Table 3. Column 1 replicates the OLS findings from Table 1, whilst columns 2-4 show the findings after instrumenting for fat mass. Controlling for all covariates, the OLS results show that fat mass negatively affects school performance. When

[22] The contextual variables *X* include: child birth weight, age in months, number of older and younger siblings under 18, log equivalised family income and its square, mother's and mother's parents' educational level, lone parent status, social class, maternal age at birth, parents' employment status, and IMD. Indicators for mother's health and behaviour *P* include: mother's smoking and drinking during pregnancy, breastfeeding, mother's 'locus of control', two measures of maternal mental health (EPDS and CCEI), parental involvement or interest in the child's development, and parents' engagement in active (outdoor) activities with their child.

[23] Girls perform better, and the child's age is positively related to performance. Test scores are lower for those brought up by a stepfather rather than the natural father. Mother's education and father's social class are positively related to the child's test score. Mother's employment status negatively affects the child's performance, with larger coefficients for full-time employment compared to part-time. Test scores are lower for those living in more deprived areas and for those whose mothers have an external locus of control. Mother's teaching and child related activity scores show positive associations with school performance.

[24] As a general test of gene-environment interactions, we explore whether our genetic variants are only expressed in specific environments, and therefore whether there is any direct evidence of violation of the monotonicity assumption. We re-estimate the first stage regression, interacting the genetic variants with indicators for various subgroups and test whether the *FTO* and *MC4R* coefficients are the same across groups. We specify the following subgroups: gender, duration of breastfeeding, social class, mother's educational level, and quartiles for birth weight, log income, the Index of Multiple Deprivation (IMD) and mother's teaching score. The results (available from the authors) show no significant differences in the estimates, suggesting that gene-environment interactions do not play an important role for the genetic variants used here.

genetic variants are used as instruments however, none of the estimates show significant effects of fat mass on educational performance.[25] The Hansen's J-test does not reject the validity of Assumption 1, i.e. there is no evidence that the genetic variants are correlated with the child's unobserved characteristics.

Although the IV estimates are of opposite sign and more than double the magnitude of the OLS estimate, the large standard errors preclude us from rejecting the null of no effect. In fact, the Hausman tests for the endogeneity of fat mass shows that the OLS and IV estimates are not significantly different. However, even if one were willing to assume that including the wide range of variables in the OLS regression solves the endogeneity problem – and thus interpreting the OLS estimate as the average causal effect – the magnitude of the estimate shows that, if any, the effect is very small. The OLS confidence interval is [-0.066, -0.014], indicating that a one standard deviation increase in adiposity leads to a maximum of 0.066 standard deviations decrease in educational outcomes. This corresponds to a decrease in children's KS3 score from the median to the 47th percentile. Hence, even if there truly were an adverse effect of adiposity on educational outcomes, the magnitude of this effect is very small.

5.3 Multivariate Imputation

Although the IV regressions show no evidence of adiposity affecting outcomes, we cannot statistically distinguish the OLS estimates from the IV due to the relative imprecision of the latter. One way of dealing with this is to use a larger sample. For all variables apart from the genetic variants, we therefore impute their missing values using multivariate imputation.[26] This leads to an almost doubling of our sample to 7,368 observations.[27] Including the full set of controls, the OLS estimate is virtually unchanged, -0.049, with $p<0.000$, and the IV estimate is now slightly smaller, 0.0851, with $p=0.407$. Despite the somewhat smaller standard errors compared to those derived using the original complete data, they remain too

[25] Appendix D presents the estimates that distinguish between using only *MC4R*, and only (specific values of) *FTO*. As discussed in section 3.1, the final IV estimate in Table 3 is a weighted average of these separate regressions.

[26] As the genetic variants do not show any systematic correlation to the other covariates in the model apart from child fat mass, we cannot impute its values; the variants are distributed randomly.

[27] Due to the strong within individual association of fat mass and of BMI, and as we observe children's fat mass, BMI and weight at various ages, we have strong predictors of child fat mass if it is missing at age 11. Similarly, a child's performance on the Key Stage tests (taken at 7 and 11 as well as 14) is highly correlated over time. As we observe the child's scores on the entry assessment test, as well as their KS exams at age 7 and 11, this will help in imputing any missing KS3 scores.

large to statistically distinguish between OLS and IV (*p=0.185*).

6. Conclusion and Discussion

The increasing availability of biomedical data, in combination with a growing medical literature on the effects of carrying specific genetic variants, introduces a different approach to the examination of certain risk factors on economic outcomes. This paper discusses the method of Mendelian randomization and undertakes an application to child adiposity and academic performance.

We discuss the specific conditions that need to be met for genetic variants to be used as instruments for identification of the average causal response. These conditions have not been well defined in the current economic literature, but the increasing availability of biomedical data makes understanding of these conditions crucial to the successful use of genotypes as instruments for modifiable risk factors.

Our empirical application uses recently identified genetic variants as instruments for adiposity. We argue that these variants are the best current candidates for use as genetic markers: they have been shown to be associated with adiposity in large population samples and we argue that they are likely to meet the conditions required for suitable instruments. We also use direct measures of body fat mass, rather than the generally used BMI. OLS shows that leaner children perform better in school tests compared to their more adipose counterparts. Our genetic IV analysis, however, shows no evidence that children's fat mass affects their academic performance, although the estimates are imprecise. While an endogeneity test does not allow us to statistically distinguish between the OLS and IV estimates, the magnitude of the estimates shows that any effect is very small. Based on our robust IV approach, we therefore conclude that adiposity is not a major determinant of educational outcomes and encourage further research to move on to examine other modifiable risk factors for low educational attainment.

Our discussion of the conditions for the suitability of genetic variants as instruments and our application raise a more general issue of the use of genetic variants as instrumental variables in economic settings. In our case, while our instruments are not weak in a statistical sense, their effects may be too small to impact on the possible pathways to academic performance. In other words, a two kilogram increase in adiposity may not lead to a large drop in self-esteem or an increase in absenteeism. To that end, it is perhaps not surprising that we do not

find a significant effect on academic performance. That said, *FTO* is the strongest adiposity-marker yet identified. It is relatively unlikely that common variants will be found with larger adiposity-effects, as those with larger effects tend to be discovered before smaller ones (though rare variants with stronger effects may be identified).

This illustrates the more general question of whether genetic variants are powerful enough to identify causal effects in studies examining economic outcomes.[28] The answer to this question will depend on the variant, the risk factor and the outcome of interest. With a rapidly growing medical literature on the effects of carrying specific genetic variants, one option is to wait for more variants to be identified and to combine these into one or more instrumental variables, such as a count of the number of risk alleles. This could increase the explained phenotypic variation and with that, the precision of the estimates. But as noted, in the case of adiposity - and another physical attributes that economists have been interested in such as height - any additional variants are likely to have even smaller effects than those already identified.

In conclusion, we argue that genetic instruments need to be used with care. Their appropriate use requires that several conditions, which have not hitherto been spelt out in economic applications, are met. But even if these conditions are met, the sample sizes in data sets that contain both genetic markers and outcomes of interest to economists may be too small to obtain definitive results. With a rapid increase in the number of genome wide association studies being done, and with a decrease in their costs, this may change, but at present we argue that caution is required until we obtain data on genetic variation in data sets containing outcome measures that are of sufficiently large size to use this approach.

[28] Numerous studies have shown the benefits of using Mendelian randomization in other settings, such as medicine and epidemiology (see e.g. Chen et al., 2008)

References

Angrist, Joshua, and Jorn-Steffen Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion* (Princeton: Princeton University Press, 2009).

Angrist, Joshua D., Kathryn Graddy, and Guido W. Imbens, "The Interpretation of Instrumental Variables Estimators in Simultaneous Equation Models with an Application to the Demand for Fish," *Review of Economic Studies* 67 (2000), 499-527.

Angrist, Joshua, and Alan B. Krueger, "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics* 106:4 (1991), 979-1014.

Averett, Susan L., and David C. Stifel, "Food for Thought: The Cognitive Effects of Childhood Malnutrition in the United States," (2007) http://aysps.gsu.edu/events/2007/averett_stifel_obesity_june.pdf

Bhatti, Parveen, Alice J. Sigurdson, Sophia S. Wang, Jinbo Chen, et al., "Genetic Variation and Willingness to Participate in Epidemiologic Research: Data from Three Studies," *Cancer Epidemiology Biomarkers and Prevention* 14 (2005), 2449-53.

Burkhauser, Richard, and John Cawley, "Beyond BMI: The Value of More Accurate Measures of Fatness and Obesity in Social Science Research," *Journal of Health Economics* 27 (2008), 519-29.

Cawley, John, "The Impact of Obesity on Wages," *Journal of Human Resources* 39:2 (2004), 451-74.

Chen, Lina, George Davey Smith, Roger M. Harbord, and Sarah J. Lewis, "Alcohol Intake and Blood Pressure: A Systematic Review Implementing a Mendelian Randomization Approach,"*PLoS Medicine* 5:3 (2008), 461-71.

Colhoun, Helen M., Paul M. McKeigue, and George Davey Smith, "Problems of Reporting Genetic Associations with Complex Outcomes," *The Lancet* 361 (2003), 865-72.

Davey Smith, George, and Shah Ebrahim, "'Mendelian Randomization': Can Genetic Epidemiology Contribute to Understanding Environmental Determinants of Disease?" *International Journal of Epidemiology* 32 (2003), 1-22.

Davey Smith George, Debbie A. Lawlor, Roger Harbord, Nic Timpson, et al., "Clustered Environments and Randomized Genes: A Fundamental Distinction between Conventional and Genetic Epidemiology," *PLoS Medicine* 4 (2008), 1985-92.

Ding, Weili, Steven F. Lehrer, J. Niels Rosenquist, and Janet Audrain-McGovern, "The Impact of Poor Health on Education: New Evidence using Genetic Markers," *Journal of Health Economics* 28:3 (2009), 578-97.

Fischer, Julia, Linda Koch, Christian Emmerling, Jeanette Vierkotten, et al., "Inactivation of the Fto Gene Protects from Obesity," *Nature* 458 (2009), 894-9.

Fletcher, Jason M., and Steven F. Lehrer, "Using Genetic Lotteries within Families to Examine the Causal Impact of Poor Health on Academic Achievement," (2008) http://www.econ.sfu.ca/Seminars/SeminarPapers/Lehrer.pdf

Frank, Richard G., and Ellen Meara, "The Effect of Maternal Depression and Substance Abuse on Child Human Capital Development," NBER Working paper 15314 (2009).

Frayling, Timothy M., Nicholas J. Timpson, Michael N. Weedon, Eleftheria Zeggini, et al., "A Common Variant in the FTO Gene is Associated with Body Mass Index and Predisposes to Childhood and Adult Obesity," *Science* 316 (2007), 889-94.

Geier, Andrew B., Gary D. Foster, Leslie G. Womble, Jackie McLaughlin, et al., "The Relationship between Relative Weight and School Attendance among Elementary Schoolchildren," *Obesity* 15:8 (2007), 2157-61.

Golding, Jean, Marcus Pembrey, Richard Jones, and the ALSPAC Study Team, "ALSPAC - The Avon Longitudinal Study of Parents and Children: I. Study Methodology," *Pediatric and Perinatal Epidemiology*, 15 (2001), 74-87.

Guo, Guang, Kari E. North, and Seulki Choi, "DRD4 Gene Variant associated with Body Mass: The National Longitudinal Study of Adolescent Health," *Human Mutation* 27:3 (2006), 236-41.

Hinney, Anke, Jennifer Schneider, Andreas Ziegler, Gerd Lehmkuhl, et al., "No Evidence for Involvement of Polymorphisms of the Dopamine D4 Receptor Gene in Anorexia Nervosa, Underweight, and Obesity," *American Journal of Medical Genetics* 88:6 (1999), 594-7.

Kaestner, Robert, and Michael Grossman, "Effects of Weight on Children's Educational Achievement," NBER Working paper 13764 (2008).

Kivimäki, Mika, George Davey Smith, Nicholas J. Timpson, Debbie A. Lawlor, G. David Batty, et al., "Lifetime Body Mass Index and Later Atherosclerosis Risk in Young Adults: Examining Causal Links using Mendelian Randomization in the Cardiovascular Risk in Young Finns Study," *European Heart Journal* 29:20 (2008), 2552-60.

Lawlor, Debbie A., Frank Windmeijer, and George Davey Smith, "Is Mendelian Randomization "Lost in Translation"?" *Statistics in Medicine* 27 (2008), 2750-55.

Lawlor, Debbie A., Roger M. Harbord, Jonathan A. Sterne, Nic Timpson, and George Davey Smith, "Mendelian Randomization: Using Genes as Instruments for Making Causal Inferences in Epidemiology," *Statistics in Medicine* 27 (2008), 1133-63.

Loos, Ruth, J.F., Cecilia M. Lindgren, Shengxu Li, Eleanor Wheeler et al., "Common

Variants Near *MC4R* are Associated with Fat Mass, Weight and Risk of Obesity," *Nature Genetics* 40:6 (2008), 768-75.

Norton, Edward C., and Euna Han, "Genetic Information, Obesity and Labor Market Outcomes," *Health Economics 17:9 (2008), 1089-1104*.

Patsopoulos, Nikolaos A., Athina Tatsioni, and John P. Ioannidis, "Claims of Sex Differences: an Empirical Assessment in Genetic Associations," *JAMA* 298 (2007): 880-93.

Petterson, Stephen M., and Alison B. Albers, "Effects of Poverty and Maternal Depression on Early Child Development," *Child Development* 72:6 (2001), 1794-1813.

Plomin, Robert, *Development, Genetics and Psychology*, (Mahwah: Lawrence Erlbaum Associates Inc., 1986).

Redline, Susan, Peter V. Tishler, Mark Schluchter, Joan Aylor, et al. "Risk Factors for Sleep-Disordered Breathing in Children: Association with Obesity, Race and Respiratory Problems," *American Journal of Respiratory and Critical Care Medicine* 159:5 (1999), 1527-32.

Sabia, Joseph J., "The Effect of Body Weight on Adolescent Academic Performance," *Southern Economic Journal* 73:4 (2007), 871-900.

Schwartz, Marlene B., and Rebecca M. Puhl, "Childhood Obesity: A Societal Problem to Solve," *Obesity Reviews* 4:1 (2003), 57-71.

Timpson, Nicholas J., Pauline M. Emmett, Timothy M. Frayling, Imogen Rogers, et al., "The Fat Mass- and Obesity-Associated Locus and Dietary Intake in Children," *The American Journal of Clinical Nutrition* 88:4 (2008), 971-8.

Timpson, Nicholas J., Roger Harbord, George Davey Smith, Jeppe Zacho, et al., "Does Greater Adiposity Increase Blood Pressure and Hypertension Risk? Mendelian Randomization Using the FTO/MC4R Genotype," *Hypertension* 54 (2009), 84-90.

Wardle, Jane, Susan Carnell, Claire M.A. Haworth, I. Sadaf Farooqi, Stephen O'Rahilly, and Robert Plomin, "Obesity Associated Genetic Variation in *FTO* is associated with Diminished Satiety," *The Journal of Clinical Endocrinology and Metabolism* 93:9 (2008), 3640-3.

Willer, Cristen J., Elizabeth K. Speliotes, Ruth J.F. Loos, Shengxu Li, Cecilia M. Lindgren, et al., "Six New Loci Associated with Body Mass Index Highlight a Neuronal Influence on Body Weight Regulation," *Nature Genetics* 41 (2008): 25-34.

Appendix A: A Brief Introduction to Genetics

Each cell in the human body contains a nucleus in which most DNA (99.9995%) is kept.[29] DNA is stored in structures called chromosomes, where each chromosome contains a single continuous piece of DNA. All cells in the human body apart from gametes (i.e. germ cells) contain 46 chromosomes, organised into 23 chromosome pairs: one copy of chromosome 1-22 from each parent, plus an X-chromosome from the mother and either an X or a Y chromosome from the father.

Locations (or loci) where DNA varies between people are called polymorphisms. The most commonly studied form of polymorphism is a Single Nucleotide Polymorphism (SNP): a single base-pair variation in a DNA locus. As chromosomes come in pairs, humans have two base-pairs at each locus, called alleles. These alleles can either be the same or different. The term genotype is used to describe the specific set of alleles inherited at a particular chromosome locus. For example, individuals can have one of three genotypes of the FTO SNP (one of the genetic variants used here): they can be homozygous for the common allele (TT), heterozygous (AT), and homozygous for the rare allele of *FTO* (AA).[30] The visible or measurable effect of a particular genotype is called the phenotype.

The phenotype we examine in this paper is child fat mass. Studies that examine the heritability of adiposity generally report large proportions of the variance that are due to genetics: between 0.4 and 0.7 (Plomin, 1986).[31] A high heritability however, does not imply that any individual genetic variant has large phenotypic effects. For example, there are many different SNPs that affect human weight, though all with small effects: so-called 'polygenes'. Together, these variants may have a large phenotypic effect.

[29] A small amount of DNA exists in the mitochondria, structures that supply the cell with energy. The remainder of this section refers only to nuclear DNA, which is the DNA used to obtain genetic variation in this (and most genetic epidemiology) studies.

[30] Conventionally, italics are used to indicate the name of a genetic variant (e.g. *FTO*); not italicising indicates the protein influenced by a particular genetic variant (e.g. the FTO SNP). We use the same convention.

[31] The heritability of a characteristic is defined as the proportion of the total variance that is explained by genetic factors. It is most commonly calculated from twin studies by comparing intra-pair correlations for a characteristic in monozygotic (MZ) twins with intra-pair correlation in dizygotic (DZ) twins. The heritability is calculated as twice the difference between MZ and DZ intra-pair correlations ($h^2 = 2*(r_{MZ}-r_{DZ})$).

Appendix B: *FTO*, *MC4R* and a Random Set of Additional Variables

Table B1. Coefficients (std err) of the indicators presented in the first column regressed on *FTO* and *MC4R*

	FTO		*MC4R*	
Sleep variables				
Length of night's sleep (school day), 81 months	0.014	(0.017)	-0.011	(0.020)
Length of night's sleep (Saturday), 81 months	0.011	(0.022)	-0.031	(0.025)
Child has difficulty sleeping, 81 months[1]	-0.016	(0.011)	0.007	(0.013)
Sleeping problem anxiety score, 81 months	0.029	(0.034)	-0.010	(0.036)
Behaviour / self-esteem child				
Child is picked on / bullied, 9 years[1]	-0.009	(0.009)	-0.015	(0.011)
Depression score child, 9 years	0.325	(0.225)	0.164	(0.258)
Anti-social score child, 9 years	0.216	(0.206)	-0.386	(0.239)
Child locus of control, 8 years	0.023	(0.244)	-0.264	(0.291)
Child's scholastic competence score, 8 years	0.063	(0.241)	-0.110	(0.283)
Child's global self worth score, 8 years	-0.012	(0.237)	-0.178	(0.279)
Child's total self esteem, 8 years	0.019	(0.205)	-0.123	(0.240)
Strength & Difficulties Questionnaire (SDQ)				
Anti-social behaviour (mother-reported), 9 years	-0.044	(0.224)	0.268	(0.256)
Hyperactive behaviour (mother-reported), 9 years	-0.262	(0.217)	0.407	(0.257)
Emotional problems (mother-reported), 9 years	-0.152	(0.218)	0.027	(0.254)
Conduct problems (mother-reported), 9 years	-0.056	(0.220)	0.305	(0.260)
Peer problems (mother-reported), 9 years	-0.300	(0.216)	0.036	(0.259)
Learning difficulties				
Freq. to special class due to learning difficulties, 81 months	0.013	(0.014)	0.021	(0.019)
Freq. to special class due to learning difficulties, 9 years	0.003	(0.020)	-0.004	(0.023)
Freq. to special class due to learning difficulties, 11 years	0.033*	(0.018)	0.022	(0.022)
Child ever had speech/language therapy, 91 months[1]	-0.004	(0.007)	0.002	(0.008)
Child has dyslexia (mother-reported)[1]	-0.002	(0.004)	0.003	(0.005)
Child is autistic (mother-reported)[1]	0.001	(0.002)	0.001	(0.002)
Mother's health and behaviour				
Mother's self-esteem (Bachman score)	0.310*	(0.184)	0.056	(0.220)
Mother's depression score, 18 weeks gestation	0.037	(0.032)	0.004	(0.038)
Mother's somatic problems score, 18 weeks gestation	-0.001	(0.035)	0.033	(0.040)
Financial situation of the household				
House is owner-occupied, 21 months[1]	-0.005	(0.007)	0.008	(0.008)
House is rented or via housing association, 21 months[1]	0.001	(0.005)	0.001	(0.006)
Council housing, 21 months[1]	0.002	(0.005)	-0.010*	(0.005)
Indicators at birth of child				
Month of birth (1=September, 12 = October)	0.016	(0.039)	0.012	(0.043)
Admission to special care birth unit[1]	-0.008	(0.005)	0.005	(0.006)
Multiple births (twins or triplets)[1]	-0.001	(0.003)	0.002	(0.004)
Gestational age at delivery	0.027	(0.032)	0.025	(0.039)
Caesarean section[1]	-0.006	(0.007)	0.008	(0.008)
Different measures of child weight / fat mass				
Waist circumference, 11 years	1.14***	(0.198)	0.85***	(0.242)
Hip circumference, 11 years	0.95**	(0.172)	0.43**	(0.210)

Notes: [1] Binary indicator. * $p<0.10$; ** $p<0.05$; *** $p<0.01$

Appendix C: Descriptive Statistics by *FTO* and *MC4R*: Mean (st. dev.)

	FTO (rs9939609)				*MC4R* (rs17782313)			
	TT	AT	AA	*p*-value	TT	CT	CC	*p*-value
Adiposity								
Fat mass, age 11	98.7	100.3	101.6	< 0.000	99.6	100.3	100.7	0.020
	(9.33)	(10.00)	(10.62)		(9.70)	(10.17)	(10.27)	
BMI, age 11	98.6	100.0	101.3	< 0.000	99.5	100.1	99.8	0.031
	(9.18)	(10.00)	(10.03)		(9.61)	(10.07)	(9.03)	
Weight (kg), age 11	42.5	43.6	45.1	< 0.000	43.3	43.6	44.1	0.169
	(9.47)	(10.02)	(10.05)		(9.64)	(10.06)	(10.74)	
Height (cm), age 11	150.5	150.5	151.1	0.114	150.7	150.4	150.5	0.325
	(7.10)	(7.17)	(7.09)		(7.03)	(7.26)	(7.38)	
Contextual variables								
Birth weight (g)	3414	3426	3429	0.518	3434	3397	3465	0.366
	(556)	(532)	(557)		(544)	(547)	(531)	
Breastfeeding	1.86	1.90	1.89	0.473	1.84	1.96	1.85	0.037
	(1.21)	(1.19)	(1.22)		(1.22)	(1.17)	(1.19)	
Age at KS3	169.60	169.62	169.68	0.677	169.55	169.68	169.97	0.113
	(3.69)	(3.77)	(3.69)		(3.72)	(3.72)	(3.78)	
Ln(income)	5.31	5.33	5.33	0.255	5.33	5.31	5.31	0.260
	(0.46)	(0.44)	(0.44)		(0.45)	(0.45)	(0.41)	
Mother's education	2.36	2.37	2.35	0.995	2.36	2.37	2.29	0.619
	(0.86)	(0.90)	(0.86)		(0.88)	(0.88)	(0.88)	
Social class	3.07	3.05	2.91	0.023	3.03	3.02	3.07	0.950
	(1.27)	(1.30)	(1.26)		(1.29)	(1.27)	(1.30)	
Mum works PT	0.40	0.40	0.40	0.941	0.40	0.41	0.39	0.841
	(0.49)	(0.49)	(0.49)		(0.49)	(0.49)	(0.49)	
Mum works FT	0.10	0.09	0.10	0.906	0.11	0.09	0.06	0.019
	(0.30)	(0.29)	(0.30)		(0.31)	(0.29)	(0.24)	
Partner employed	0.95	0.96	0.94	0.944	0.96	0.94	0.94	0.180
	(0.34)	(0.32)	(0.35)		(0.34)	(0.32)	(0.30)	
IMD	18.81	19.57	19.59	0.154	19.53	19.01	18.84	0.234
	(13.78)	(14.07)	(13.84)		(14.15)	(13.73)	(12.97)	
Mother's health and behaviour								
Alcohol	0.55	0.58	0.56	0.309	0.57	0.56	0.58	0.770
	(0.50)	(0.49)	(0.50)		(0.50)	(0.50)	(0.49)	
Smoke	0.18	0.18	0.15	0.400	0.17	0.18	0.17	0.743
	(0.38)	(0.39)	(0.36)		(0.38)	(0.39)	(0.38)	
Mother's locus of control	98.49	98.95	99.44	0.035	98.69	99.03	99.47	0.158
	(9.36)	(9.56)	(9.58)		(9.62)	(9.35)	(9.18)	
EPDS	6.27	6.46	6.30	0.610	6.33	6.44	6.24	0.719
	(4.58)	(4.57)	(4.44)		(4.64)	(4.42)	(4.53)	
CCEI	12.27	12.64	12.36	0.507	12.30	12.72	12.42	0.183
	(7.26)	(7.26)	(6.98)		(7.20)	(7.29)	(6.87)	
Teaching score	7.01	7.03	7.07	0.195	7.06	6.99	7.01	0.058
	(0.93)	(0.93)	(0.86)		(0.88)	(0.96)	(0.95)	
Activities (indoor)	0.69	0.68	0.69	0.623	0.68	0.69	0.70	0.153
	(0.20)	(0.21)	(0.20)		(0.21)	(0.20)	(0.18)	
Activities (outdoor)	27.81	27.87	27.96	0.481	27.85	27.96	27.25	0.571
	(4.67)	(4.66)	(4.31)		(4.55)	(4.62)	(5.14)	
N	1356	1784	589		2140	1390	199	

Notes: *p*-values correspond to the coefficient of an OLS regression of the variable in column 1 on the variable indicating the number of *FTO* or *MC4R* risk alleles.

Appendix D: IV Estimation using Different Instrument Sets

Table D1 presents the IV results using different instrument sets. As discussed in section 3.1, our IV estimates are weighted averages of (a) the Wald estimators using the two instruments *FTO* and *MC4R*, and (b) the average causal derivatives calculated at the two values of *FTO*, with the weights proportional to the adiposity change induced by the instrument, i.e. the instrument strength in the first stage.

Here we show the IV estimates using different combinations of instruments. Column (1) replicates our IV estimate as shown in Table 3, using the linear indicator for *FTO* and the binary *MC4R*. Column 2 shows the IV estimate when using only the binary *MC4R*; column 3 only uses the linear *FTO*. Columns 4 and 5 distinguish between the two binary indicators for *FTO*, and column 6 includes the two dummies simultaneously.

Using the *MC4R* indicator alone, the IV estimate is 0.035; using the corresponding (linear) *FTO* results in an estimate of 0.149. The final estimate of 0.137 (column 1) is closer to that using *FTO* (column 3) than *MC4R* (column 2), reflecting the higher weights due to the larger adiposity change induced by *FTO*. Distinguishing between the two values of *FTO*, the IV estimate for the (0/1) and (1/2) indicators are -0.084 and 0.580 respectively. Although this difference seems large, the standard errors are also very large. Once both *FTO* genotypes are included together – either linearly (0/1/2, column 3) or as two binary indicators as in column 6 – the IV estimate is very similar to the final estimate as shown in Table 3, confirming that – also with different instrument combinations – we find no evidence of a causal effect of adiposity on educational outcomes.

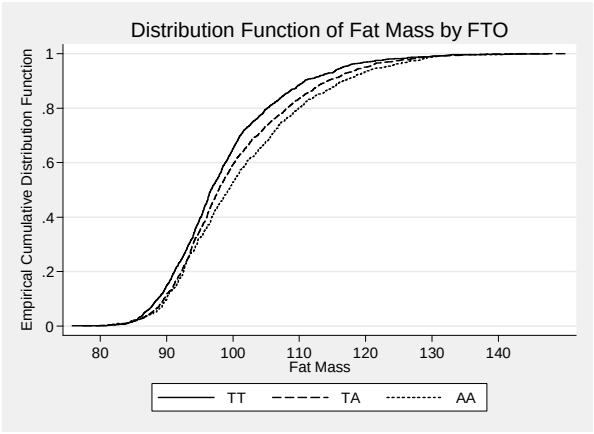
Table D1. IV estimates of the average response in standardised KS3 to a 1 standard deviation change in fat mass, different instrument sets

	(1) *MC4R* (0/1), *FTO* (0/1/2)	(2) *MC4R*, binary(0/1)	(3) *FTO*, Linear(0/1/2)	(4) *FTO*, binary (0/1)	(5) *FTO*, binary (1/2)	(6) *FTO*, Two binary indicators
Fat Mass	0.137	0.035	0.149	-0.084	0.580	0.149
	p=0.302	*p=0.928*	*p=0.289*	*p=0.673*	*p=0.143*	*p=0.287*
	[-0.12, 0.40]	[-0.72, 0.79]	[-0.13, 0.42]	[-0.47, 0.31]	[-0.20, 1.36]	[-0.13, 0.42]
N	3729	3729	3729	3140	2373	3729
Controls	No	No	No	No	No	No

Notes: Column 1 is the specification as in Table 3, not controlling for any covariates: using the binary *MC4R* and linear *FTO*. Column 2 only uses *MC4R*; Column 3 only uses the linear *FTO*. Column 4 and 5 include *FTO* (0/1) and *FTO* (1/2) respectively. Column 6 specifies both binary *FTO* indicators as instruments. 95% confidence intervals in square brackets; *p* is p-value for standard t-ratio

Figures

Figure 1a: Distribution and Weight Functions of Fat Mass for *FTO* genotypes



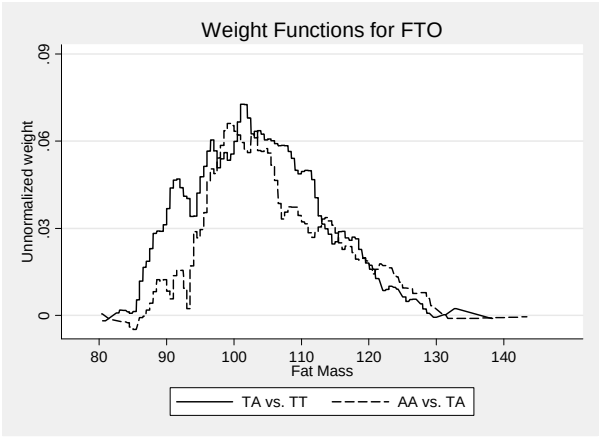
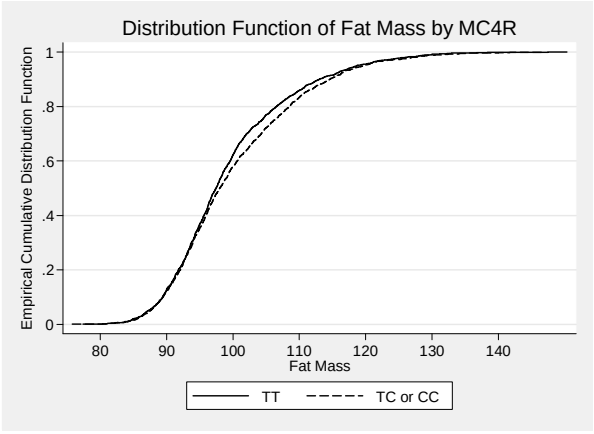
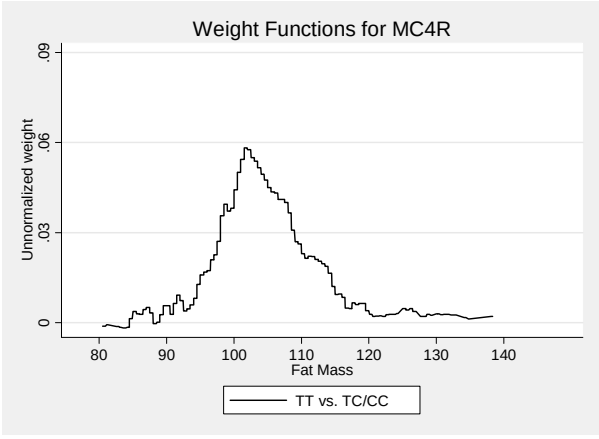


Figure 1b: Distribution and Weight Functions of Fat Mass for *MC4R* genotypes





Tables

Table 1. OLS estimates of the average response in standardised KS3 to a 1 standard deviation change in fat mass

	(1)	(2)	(3)
Fat Mass	-0.099	-0.052	-0.040
	p<0.000	*p<0.000*	*p=0.002*
	[-0.128, -0.070]	[-0.078, -0.026]	[-0.066, -0.014]
R-squared	0.01	0.26	0.30
Number of children	3729	3729	3729
Included control variables:			
Contextual variables	No	Yes	Yes
Mother's health and behaviour	No	No	Yes

Notes: 95% confidence intervals in square brackets; *p* is p-value for standard t-ratio. Footnote 22 lists all covariates.

Table 2. First stage OLS estimates of the average response in fat mass per 1 unit change in *FTO* and *MC4R*

	(1) Fat Mass	(2) Fat Mass	(3) Fat Mass
FTO	1.491	1.404	1.432
	p<0.000	*p<0.000*	*p<0.000*
	[1.028, 1.954]	[0.951, 1.858]	[0.981, 1.883]
MC4R	0.768	0.890	0.898
	p=0.019	*p=0.006*	*p=0.005*
	[0.124, 1.412]	[0.261, 1.519]	[0.269, 1.527]
IV strength, F-statistic	22.8	22.7	23.5
R-squared	0.01	0.07	0.08
Number of observations	3729	3729	3729
Included control variables:			
Contextual variables	No	Yes	Yes
Mother's health and behaviour	No	No	Yes

Notes: 95% confidence intervals in square brackets; *p* is p-value for standard t-ratio. See footnote 22 for list of covariates.

Table 3. IV estimates of the average response in standardised KS3 to a 1 standard deviation change in fat mass

	OLS	IV		
	(1)	(2)	(3)	(4)
Fat Mass	-0.040	0.137	0.098	0.115
	p=0.002	*p=0.302*	*p=0.408*	*p=0.312*
	[-0.066, -0.014]	[-0.123, 0.396]	[-0.133, 0.328]	[-0.108, 0.337]
p-value, Hansen J-test		0.783	0.811	0.923
p-value, Hausman test		0.062	0.193	0.161
Number of children	3729	3729	3729	3729
Included control variables:				
Contextual variables	Yes	No	Yes	Yes
Mother's health & behaviour	Yes	No	No	Yes

Notes: 95% confidence intervals in square brackets; *p* is p-value for standard t-ratio. See footnote 22 for list of covariates.