

Child Height, Health and Human Capital: Evidence using Genetic Markers

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

20 September 2010

Height has long been recognised as associated with better outcomes: the question is whether this association is causal. We use children's genetic variants as instrumental variables (IV) to deal with possible unobserved confounders and examine the effect of child and adolescent height on a wide range of outcomes: academic performance, IQ, self-esteem, symptoms related to depression and behavioural problems, including hyperactivity, emotional, conduct and peer problems. OLS findings show that taller children have higher IQ scores, perform better in school tests, and are less likely to have emotional or peer problems. The IV results differ. They show that taller children have better cognitive performance but, in contrast to the OLS, indicate that taller children are more likely to have behavioural problems. The magnitude of these IV estimates is large. For example, the effect of one standard deviation increase in height on IQ is comparable to the IQ difference for children born approximately 6 months apart within the same school year, while the increase in hyperactivity is comparable to the raw difference in hyperactivity between boys and girls.

Key words: Child and Adolescent Height; Human Capital; Mental Health; Behavioural Outcomes; Instrumental Variables; Mendelian Randomization; Genetic Variants; ALSPAC

JEL: I1, J24

Acknowledgements:

We thank John Cawley, George Leckie, Sarah Smith, and Nicholas Timpson 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.

Funding:

The UK Medical Research Council (MRC), the Wellcome Trust and the University of Bristol provide core support for ALSPAC. GDS and DAL work in a centre that receives funding from the UK MRC (G0600705) and University of Bristol. 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). No funding body influenced data collection, analysis or its interpretation. This publication is the work of the authors, who will serve as guarantors for the contents of this paper.

¹Imperial College Business School, Imperial College London; CMPO, University of Bristol

²MRC Centre for Causal Analyses in Translational Epidemiology (CAiTE), School of Social and Community Medicine, University of Bristol

³CMPO and Department of Economics, University of Bristol; Imperial College Business School, Imperial College London; CEPR

⁴CMPO and Department of Economics, University of Bristol; Centre for Microdata, Methods and Practice

1. Introduction

The association between height and wealth has been noted in the academic literature for many decades. As early as the 17th Century, Guarinoni – one of the founders of preventive medicine – pointed to the difference in growth rates between the rich in towns and the poor in the countryside (Tanner, 1982). More recent studies find height to be positively related to education (Magnusson, Rasmussen and Gyllensten, 2006) and income (Persico, Postlewaite and Silverman, 2004). The advantages associated with greater height have also been reported for children. For example, Case and Paxson (2008) find that taller children perform better in school tests compared to shorter children and suggest that the relationship between childhood height and income and education in adulthood is due to height being associated with greater intelligence.

One problem in estimating the relationship between height and outcomes is that the relationship may not be causal. Height is influenced by a wide range of environmental factors experienced in childhood which may be the determinants of the outcomes, rather than height per se, for example, unobserved family wealth or differences in children's nutrition. To the extent that these unobserved differences are family specific, one approach is to identify the causal impact from twin or sibling differences in height and outcomes. Case and Paxson (2010) use this approach, exploiting differences between siblings. They conclude that taller children perform better in school, progress faster through school and consider themselves more scholastically competent than their shorter siblings.

However, accounting for fixed unobserved family effects using twin (or sibling) differences does not necessarily eliminate the inconsistency of the conventional cross-sectional estimator and can even aggravate it (Griliches, 1979; Bound and Solon, 1999). The intuition is that taking twin or sibling differences filters out some, but not all, endogenous variation but also filters out exogenous variation. If the endogenous variation comprises as large a proportion of the remaining within-sibling variation as it does of the between-sibling variation, the parameters using within-sibling estimation are as vulnerable to endogeneity bias as that found in between-sibling estimation. The inconsistency of the within-sibling estimator is less than the between-sibling estimator only if the endogenous variation comprises a smaller share of the within-sibling variation in height than it does of the between-sibling variation (Bound and Solon, 1999). There is no reason to be confident that this is the case.

This paper therefore takes a different approach to estimate the causal effect of child height on children's cognitive and non-cognitive outcomes. We exploit differences in children's genetic make-

up as instrumental variables (IV) for their height.¹ At conception, genes are randomly allocated from parents 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). Furthermore, since genetic variation is determined at conception, it cannot be affected by later outcomes. Hence, in addition to dealing with reverse causality and fixed characteristics that affect both height and the outcome, Mendelian randomization can also deal with *time-varying* characteristics that affect height and outcomes. Therefore, under certain assumptions that we discuss below, genetic variants will allow us to isolate the causal effect of child height on the outcome of interest.

This paper is the first to exploit genetic variants for height in an attempt to estimate the causal effect of height on cognitive and non-cognitive outcomes for children. We begin therefore by outlining the conditions needed to use genetic variants as instruments. To examine and test the validity of the IV approach in our context, we show first that the genetic variants are uncorrelated with a large set of family background variables which may confound the relationship between height and outcomes. We then run a ‘falsification check’ in which we investigate the effect of height on body weight. As weight and height are causally related, we should find strong positive effects in both the OLS and IV models. Any substantially different or negative effects in the IV estimates would cast doubt on the IV specification and/or the instruments used. Finding no significant differences between the OLS and IV, we then use the genetic variants as instruments to examine the relationship between height and an extensive set of cognitive, mental health and behavioural outcomes. In so doing, we add to the range of outcomes examined in the previous literature. In addition to children’s academic attainment, scholastic competence and self-worth studied by Case and Paxson (2010), we investigate the effects of height on IQ, symptoms of depression and behavioural problems, including hyperactivity, emotional, conduct and peer problems.

We use data from a cohort of UK children currently in their late teens (the ALSPAC survey, described below). The OLS results show that taller children perform better in school tests, have higher IQ, and are less likely to have emotional and peer problems, though these relationships differ slightly by gender. Tall girls have higher depression scores, but we find no evidence of differences in self-esteem for children of different heights. The IV results suggest there is a causal relationship between height

¹ This approach is also known as Mendelian randomization, see e.g. Davey Smith and Ebrahim (2003). It 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 (Davey Smith and Ebrahim, 2003). For a brief overview of genetics and some of the terms used here, see Appendix A.

and IQ, though this is not reflected in test scores. However, in contrast to Case and Paxson (2010), we find no evidence that height explains variation in scholastic self-esteem, global self-worth or depression. Further, we find evidence that height confers disadvantage rather than advantage as it *increases* hyperactive behaviour, emotional and peer problems.

The next section begins by examining the possible mechanisms through which height may be related to the outcomes of interest. In section three we set out our methodology and section four describes the data. The results are presented in section five; section six concludes.

2. Mechanisms

We examine a large set of outcomes: academic attainment, IQ, self-esteem, depression symptoms, and behavioural problems. There are two ways in which height may be related to these outcomes. First, being tall could cause differences in the outcome of interest. We define ‘causal’ however, not necessarily as height per se affecting the outcome, but as height triggering social reactions that in turn affect the outcome. Hence, we hypothesize the effect of height to run via different pathways, which we discuss below. We represent this causal effect in a Directed Acyclic Graph (DAG) in Figure 1, where child height H_i affects the outcome C_i via various pathways P rather than it affecting C_i directly (which would be indicated by the dashed directed edge).² Second, instead of there being a causal relationship, the association between height and the outcome of interest may be driven by other unobserved factors that affect both. In Figure 1, this is shown by the unobservables e_i that affect both H_i and C_i .³

Several pathways through which height can causally affect outcomes are discussed in the literature, including taller people enjoying social dominance (Hensley, 1993) and having higher self-esteem (Judge and Cable, 2004). The (sociological and psychological) literature posits several theories as to why (physical) characteristics may affect behaviour or achievement. First, the possession of certain characteristics (like being tall) can trigger expectations from others (like peers or teachers). These expectations may influence their behaviour towards the ‘possessor’, which in turn affects the possessor’s behaviour, often confirming the expectations. This self-fulfilling prophecy is also referred to as the ‘expectancy effect’ (see for example Darley and Fazio, 1980). For example, some evidence

² In the DAG, each node represents a variable, with square nodes being observed and circular nodes being unobserved.

³ In theory, poor outcomes could lead children to change their eating patterns, which may affect their growth resulting in reverse causation. We argue however, that this is very unlikely and consider this to be less of an issue than (for example) in the case of body weight. However, even if there is reverse causation, our identification strategy would deal with this.

suggests that taller people are perceived as more attractive (Macintyre and West, 1991). Attractiveness can in turn influence the behaviour and assessment of teachers (Clifford and Walster, 1973) or potential employers (Dipboye et al., 1975), causing taller children to behave and perform differently.

Second, short children are believed to have negative social experiences, including bullying, less social acceptance, and fewer friends (Sandberg and Voss, 2002; Voss and Mulligan, 2000),⁴ though it is worth noting that tallness in girls has also been shown to have similar negative psychological effects (Pyett et al., 2005; Binder et al., 1997). Having problematic social relationships can in turn affect self-esteem, social adjustment, behaviour, and scholastic performance (Morison and Masten, 1991; Parker and Asher, 1987; Wentzel, 2009). Related to this is the question of whether parents compensate or reinforce children's endowments. The former may mean that parents spend relatively more time with a small compared to a tall child, to compensate for the potential negative experiences related to short stature. As the child develops through childhood, this additional attention and support can in turn increase their cognitive skills, or reduce their behavioural problems.⁵

Another strand of the literature suggests that individuals (peers, parents, teachers as well as medical personnel) treat children at a 'size-appropriate' rather than 'age-appropriate' level: tall children are generally perceived to be (and treated as) older, whereas smaller children are treated as younger (Jones and Bayley, 1950; Rotnem et al., 1977; Underwood, 1991; Sandberg et al., 2004). Adults in turn may have different expectations depending on children's heights (Skuse et al., 1994), which can subsequently affect children's behaviour. Children who 'look young' according to their peers are perceived to be less (physically and verbally) aggressive and more emotional and passive (Sandberg et al., 2004). In addition, the literature has found taller children to have more behavioural problems, such as aggression or violent behaviour. Raine et al. (1998) find that height in 3-year-old children is associated with increased aggressiveness at age 11, and Farrington (1989) find that height at age 8-10 years is associated with violence at age 16-18 years. They argue that their early life may have taught them that it is an effective strategy in winning social conflicts, reinforcing this behaviour. In

⁴ In fact, the treatment of short children with growth hormone is, in part, based on the belief that being taller will improve short children's peer relationships (Sandberg et al., 2004; Sandberg and Voss, 2002). However, the evidence showing that short children have more negative experiences is ambiguous: a review by Sandberg and Voss (2002) for example, concludes that the psychological adaptation of shorter-than-average individuals is largely indistinguishable from others, whether in childhood, adolescence or adulthood. Others however, argue that short people may simply be discriminated against (Magnusson, Rasmussen and Gyllensten, 2006; West, 1991). Persico, Postlewaite and Silverman (2004) instead suggest that taller adolescents are more likely to participate in social activities that develop human capital. Hamermesh and Biddle (1994) also argue that (labour market) discrimination does not arise from correlations with height. Instead, they claim it is mainly based on the employee's looks ('beauty').

⁵ The evidence on whether parents compensate or reinforce children's endowments is mixed, see e.g. Griliches (1979) and Behrman et al. (1994).

contrast, smaller and physically weaker children lack the physical capacity to execute this behaviour (Raine et al., 1998).

As opposed to a causal effect, there may be other factors (represented by the unobservables e_i in Figure 1) that relate to both height and the outcome of interest and that drive the associations. One set of candidates is the pre- and postnatal environment. Regarding the latter, the fastest growth in children occurs up to age 2. There is evidence of links between early (post-natal) nutrition and child height, and between nutrition and cognitive and social development.⁶ But although early nutrition is a possible candidate, several studies have shown that even under conditions of severe malnutrition (prenatal, such as foetuses subjected to war-time famine and postnatal, such as starvation in the early years of life) complete equality in height with siblings or peers is attained before puberty (Tanner, 1978). In terms of the pre-natal diet, there is evidence that nutrition *in utero* plays an important role in child development. But nutriments which help development (such as the omega 3 fatty acids in fish and seafood consumption) may also hurt development. For example, fish and seafood are the primary source of (non-occupational) mercury exposure (Oken and Bellinger, 2008) and several studies have shown prenatal methylmercury exposure to be associated with decreased IQ and test scores (Axelrad et al, 2007; Cohen et al., 2005). Likewise, some studies find that maternal alcohol consumption and smoking during pregnancy negatively affect birth weight and child growth (Mills et al., 1984; Gilman, Gardener and Buka, 2008). Lower birth weights in turn are associated with poorer cognitive performance (Richards et al., 2002; Ericson and Kallen, 1998) and behavioural development (Elgen et al., 2002), though the literature suggests that this relationship is driven by family background characteristics rather than a specific intrauterine effect (Yang et al., 2008).⁷

This discussion suggests that the use of both OLS and mother fixed effects to estimate a causal effect can result in bias that can go in either direction. If parents compensate for their child's endowment by spending more time with shorter children, this would increase short children's performance relative to taller children, and hence including mother fixed effects would lead to the estimates being a lower bound. Similarly, if a well-balanced diet or the family's socio-economic position positively affects height, but also leads to fewer behavioural problems, the OLS is likely to under-estimate the true effect of height on behavioural problems. If, however, this same diet leads to better educational outcomes, OLS is likely to over-estimate the true effect on education. However, if certain dietary components lead to decreased cognitive functioning, the OLS may under-estimate the true effect on

⁶ For example, iron-deficiency in infants and children is associated with poorer cognitive, motor and socio-emotional function (see e.g. Lozoff et al., 2006). In addition, some studies report that iron supplementation positively affects height (Angeles et al., 1993).

⁷ There is mixed evidence on the effects of maternal smoking and alcohol consumption during pregnancy on child outcomes, with some arguing it lowers outcomes and others finding no effect (see for example Olds et al., 1994; Gilman, Gardener and Buka 2008; Kafouri et al., 2009; Davey Smith, 2008; Nilsson, 2008; Russell, 1991).

educational outcomes and IQ. Under the assumptions we discuss in detail below, the use of the child's genetic markers as instrumental variables will shed more light on these issues and will allow us to estimate the causal effect of child height.

3. Methodology

3.1. The Child Human Capital Production Function

We examine the impact of child height on three sets of outcomes: (1) cognitive skills, (2) mental health, and (3) behavioural problems.⁸ We discuss the outcomes in more detail below. As both height and outcomes differ by gender, we allow for differential effects by interacting height with gender in all analyses. We model the relationship between height and outcomes in terms of a human capital production function:

$$C_i = f(H_i, X_i, e_i), \quad (1)$$

where C_i , the outcome of interest for child i , is a function of child height (H_i , measured contemporaneously) and a set of child and family background characteristics (X_i). e_i represents any unobserved confounders. We begin with OLS using a linear version of (1):

$$C_i = \beta_0 + \beta_1 H_i + \beta_2 X_i + e_i. \quad (2)$$

The parameter of interest, the relationship between child height and the outcome variable, is β_1 . The possible endogeneity of height is characterised by the fact that the unobservable confounders e_i determine the outcome of interest C_i , but also determine height H_i , leading to biased OLS estimates. We use IV to deal with this, specifying the child's genetic variants Z_i as instruments for height. These variants are associated with H_i , and we assume that they are only associated with C_i indirectly through their association with H_i . In the absence of a constant treatment effect, we

⁸ Although available in our data, we do not examine physical health, as the IV assumptions are more likely to be problematic: genetic variants that affect height may also affect physical health. For example, individuals with higher levels of GDF5 on average have increased bone and cartilage growth. The latter in turn is protective against osteoarthritis (Sanna et al., 2008). Similarly, *HMG2* is associated with height, but is also over-abundant in many types of cancerous tumours (Ligon et al., 2005; Weedon et al., 2007), suggesting that some variants of genes that encourage growth may also cause uncontrolled cell growth in tissues leading to cancer (McEvoy and Visscher, 2009).

identify the average causal response using the standard linear IV estimator in (2) within a potential outcomes framework, following Angrist, Graddy and Imbens (2000). We briefly summarize their assumptions here.

Let C , H and Z denote random variables representing, respectively, the outcome of interest, child height and the genetic variant as IV. For simplicity, we discuss the case of a binary instrument, though in the estimation we use multiple instruments as is more common in Mendelian randomization experiments. Let $C_i(h, z)$ be the potential outcome for individual i that would be obtained if height, the treatment variable, was set to h and the instrument set to z . Equivalently, let $H_i(z)$ be the potential height for individual i when the instrument is set equal to z . We make the following assumptions:

Assumption 1. (Independence and Exclusion)

$$Z_i \perp \{C_i(h, z), H_i(z)\}_{h,z}$$

$$C_i(h, 1) = C_i(h, 0), \text{ for all } h.$$

Independence implies that the instrument is independent of the potential outcome and the potential height, for all values of h and z . In other words, the instrument is as good as randomly assigned. This is reflected by the missing edge between Z and e in the DAG in Figure 1. The missing edge between Z and C reflects the conditional independence assumption implied by the model for C , H and e : $C \perp Z | (H, e)$, i.e. the entire effect of Z on C is mediated through its association with H . Exclusion implies that the potential outcomes, at any height h , are unchanged by the presence or absence of the genetic variant.

Assumption 2. (Nonzero effect of instrument on height)

$$E[H_i(1) - H_i(0)] \neq 0$$

This implies that expected potential height is affected by the genetic variant and that therefore the coefficient in the (first stage) regression of H_i on Z_i is non-zero. Assumption 2 is reflected by the edge between G and H in Figure 1.

Assumption 3. (Monotonicity)

$$P[H_i(1) \geq H_i(0)] = 1$$

This means that the potential height for individual i with the genetic variant is at least as high as the potential height for the same individual without the genetic variant.

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

$$\begin{aligned}\hat{\beta}_{IV} &= \frac{E[C_i | Z_i = 1] - E[C_i | Z_i = 0]}{E[H_i | Z_i = 1] - E[H_i | Z_i = 0]} \\ &= \frac{\int E[g'_i(q) | H_i(0) < q \leq H_i(1)] P\{H_i(0) < q \leq H_i(1)\} dq}{\int P\{H_i(0) < q \leq H_i(1)\} dq}\end{aligned}$$

where $g'_i(q)$ is the derivative of $g_i(h)$ w.r.t. h 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

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

then

$$\hat{\beta}_{IV} = \frac{E[\beta_i \{H_i(1) - H_i(0)\}]}{E[H_i(1) - H_i(0)]}, \quad (3)$$

i.e. the IV estimate is a weighted average of the random coefficients β_i , with the weights proportional to the height change induced by the genetic variant.⁹

3.2 The Genetic Variants

We use a set of nine genetic variants (SNPs) that have all been robustly associated with height: *HMGA2* (rs1042725), *ZBTB38* (rs6440003), *GDF5* (rs6060373), *LOC387103* (rs4549631), *EFEMP1* (rs3791675), *SCMH1* (rs6686842), *ADAMTSL3* (rs10906982), *DYM* (rs8099594) and *C6orf106*

⁹ This easily extends to the case of multiple and multi-valued (rather than binary) instruments, as discussed in Angrist and Pischke (2009) and Angrist, Graddy and Imbens (2000). For a more detailed discussion of the use of genetic markers as instrumental variables from an economic perspective using a similar framework as the above, see von Hinke Kessler Scholder et al. (2010). Lawlor et al. (2008) includes a more general discussion of the situations and (biological) processes that may invalidate Mendelian randomization studies.

(rs2814993) (Weedon et al., 2007, 2008; Lettre et al., 2008; Gudbjartsson et al., 2008).¹⁰ This prior knowledge, and the fact that these associations have been replicated in different independent samples, justify the use of these variants and their compliance with Assumption 2. In section 5.2 we examine this assumption using the standard statistical tests. Note however, that although the relationships between the SNPs we use and height are robust, their phenotypic effects are small: 20 SNPs, which include the nine that we use only explain up to 3% of the variation in height (Weedon et al. 2008). We therefore combine the different SNPs into a count of the number of ‘tall’ alleles carried by each child to get around the problem of low power (see section 4.4 below for details).

Assumption 1, independence and exclusion, is not directly testable. Based on the science literature however, we argue that the instruments are valid and so that Assumption 1 holds. We also examine its validity indirectly. Violations of this Assumption can take several forms.

First, genetic confounding may occur because of population stratification, for example, due to ethnicity. This would imply that there is a systematic relationship between the allele frequency and the outcome in different ethnic groups, resulting in an association between the two at the population level but without a causal effect, violating Assumption 1. This is unlikely to affect our sample, as our cohort is recruited from a specific geographically defined region with a predominantly white population, and in our analysis, we only examine children whose mothers describe themselves and the child’s father as white.

Second, if the variants have multiple functions (known as pleiotropy), Assumption 1 could be violated if – over and above the association with height – any additional functions directly influence the outcome of interest. Similarly, if a variant is co-inherited with another genetic variant (known as being in linkage disequilibrium (LD)), violation of Assumption 1 depends on the effect of the co-inherited variant on the outcome of interest. The current evidence suggests that some height variants may indeed be pleiotropic or in LD with other variants. For example, individuals with higher levels of GDF5 on average have both increased bone and cartilage growth (Sanna et al., 2008). However, there is currently no evidence that the variants used here also affect (or are in LD with variants that affect) our outcomes of interest.¹¹

¹⁰ For example, Weedon et al. (2007) identify a common variant of the *HMG2* gene, using 4,921 individuals of European ancestry. To validate the robustness of this finding, they genotyped an additional 29,098 individuals from five further studies, including 6,827 children from the age of 7. Their findings show each copy of the height-increasing allele to be associated with an increase in height of 0.4 cm, with no differences between males and females. Similarly, Weedon et al. (2008) use data from a total of 30,147 individuals of European ancestry and identify 20 loci that robustly affect stature, including all those used here. These have since been confirmed in more independent samples (see e.g. Lettre et al., 2008; Gudbjartsson et al., 2008).

¹¹ Some literature suggests that part of the height-intelligence association is driven by a genetic component (Sundet et al., 2005), though others find no evidence of this genetic component in the height-education association (Silventoinen et al., 2000). Similarly, Magnusson et al. (2006), comparing first and second born biological brothers in Sweden, find that the taller

Although (in common with all IV approaches) we cannot test Assumption 1 directly, we can do so indirectly. In section 4.4 we examine the relationship between the genetic variants and a large set of child and family background characteristics. We find no association, suggesting that the genetic variants do not affect these potential confounders.

However, even after examining this broad range of background variables, 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 more general empirical evidence that shows that genetic variants are unlikely to be related to unmeasured confounders (see for example, Bhatti et al., 2005; Davey Smith et al., 2008; Kivimäki et al., 2008; Lawlor et al., 2008).

Given random allocation of genetic variants and the fact that individuals do not know their genotypes, we assume that an individual who carries a ‘tall’ allele is at least as tall as the same individual, had she not carried the ‘tall’ allele, thus satisfying the monotonicity Assumption 3.¹²

4. Data

We use data from a cohort of children born in the Avon area of England. Avon has approximately 1 million inhabitants, including 0.5 million in its main city, Bristol. 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 is broadly representative of the UK, though mothers were slightly more affluent compared to the general population (Golding et al., 2001).¹³

Detailed information on the children and their families has been collected from a variety of sources,

brother is significantly more likely to attend higher education. However, the height effect estimated between brothers is almost identical to that across all men, suggesting that the correlation between height and intelligence is not driven solely by genetic or environmental factors common to brothers. However, even if some genetic component affects both height and the outcome of interest, it does not necessarily imply that a specific genetic variant causes this through (for example) pleiotropy or LD. It may be caused, for example, by interactions between genes (from simple pairs to complex networks), other variation in DNA, such as Copy Number Variants (CNV, where whole sections of DNA are duplicated or deleted), and so on. To date, about 50 genes and regions of the genome have been associated with human height, explaining only 5% of its total variation. Hundreds, maybe thousands more effects are still lost in the genome (McEvoy and Visscher, 2009). Yang et al. (2010) show that the reason why most of these have not yet been detected, is because the individual SNP effects are too small to pass the stringent significance tests. Hence, it is possible for one (or more) of the nine instruments used here to be pleiotropic or in LD with a variant that directly affects our outcome. Based on the best available evidence however, we assume this is not the case and that assumption 1 holds.

¹² As this relies on knowing each individual’s counterfactual this remains an assumption. The literature only shows that at a group or population level those who possess the genetic variant are taller than those who do not. The monotonicity assumption could, for example, be violated in the presence of gene-environment interactions.

¹³ See www.bris.ac.uk/alspac for a more detailed description of the sample, its enrolment, and response rates.

including self-completed questionnaires, data extraction from medical and educational records, in-depth interviews, and clinical assessments and so our data set contains a large range of child health and development, 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 all nine genotypes, leaving us with approximately 7,100 children. Second, we drop children for whom we do not observe their height. Children were invited to attend specially designed clinics, where their anthropometric measures were recorded. As not all children attended these clinics, our sample sizes reduce to between 4,594 (age 8) and 3,867 (age 13). Finally, we restrict the sample to those children for whom we observe the outcome of interest, leading to a final sample size of around 3,900 at age 8 and 3,300 at age 13. We deal with missing values on other covariates by using multivariate imputation (Royston, 2004).

4.1. Outcome Measures

We examine three sets of outcomes. First, we observe two measures of cognitive function. These are the child's score on the nationally set Key Stage 3 (KS3) exam (taken by all 14-year-olds educated in the state sector) and the child's IQ, measured as age 8.¹⁴ Both measures are objective and comparable across all children. Increasing scores indicate better performance. It is important to note that IQ does not only measure 'innate' ability. Instead, our measure of IQ (WISC-III) is an index of general intellectual functioning, which is shaped by both inherited and acquired attributes, including any family and environmental influences. For example, there is evidence of differences in IQ between children of different quality home environments and socio-economic position (see e.g. Molfese et al. (1997) and references therein), so suggesting that the measure of IQ is not innate but rather reflects a large set of family and environmental influences that also affect children's development.

Second, we examine three measures of mental health or self-esteem: depression symptoms, scholastic competence and global self-worth. The latter two are measured at age 8, using the Harter's Self Perception Profile for Children (Harter, 1985), with increasing scores indicating higher self-esteem. The depression score is self-reported by the teenager at age 13 using the Moods and Feelings Questionnaire (Angold et al., 1995). Increasing scores indicate more depression symptoms.

¹⁴ The KS3 scores are averaged over three subjects (English, maths and science) and obtained from the National Pupil Database, a census of all pupils in England within the state school system (this includes 93% of English children), which is matched into ALSPAC. IQ is measured using the Wechsler Intelligence Scale for Children (WISC-III; Wechsler et al., 1992) and is administered by the ALSPAC psychology team.

Third, we examine the child's behavioural problems, as measured by the mother's report on the Strength and Difficulties Questionnaire (SDQ; Goodman, 1997) administered at age 13. SDQ has four sub-scores, which we examine separately (as is common in the literature). These are hyperactivity, emotional problems, conduct problems and peer problems. Increasing scores indicate increasing problems. For comparability, all outcomes are standardised on the full sample of children for whom data is available, with mean 100, standard deviation 10.

4.2. Measures of Child Height and the Genetic Variants

We examine the effect of contemporaneous height on each outcome. Height is adjusted for the exact age in month at which it is measured and standardised to have mean 100, standard deviation 10. All measurements are taken by trained nurses. We instrument height with a set of SNPs that have been consistently shown to relate to stature. These are: *HMG2*, *ZBTB38*, *GDF5*, *LOC387103*, *EFEMP1*, *SCMH1*, *ADAMTSL3*, *DYM* and *C6orf106*.¹⁵ We combine these into a count of the number of 'tall' alleles carried by each child (see section 4.4 for more details).

4.3. Controls

We observe a rich set of child and family background characteristics that we include as covariates. We control for the child's birth weight and for the number of older and younger siblings under 18 in the household. As the outcomes of interest may vary with within-year-age, we also account for the child's age (in months) at the time the outcome is measured. We control for the family's socio-economic position with various measures: log equivalised family income and its square, four binary variables for mother's and father's educational level, the mother's parents' educational level, an indicator for whether the child is raised by the natural father, variables indicating the family's social class, and parents' employment status when the child is 21 months. As a further measure, we include a measure of small (local) area deprivation, as measured at the child's birth.¹⁶

In addition to these generally observed controls, our data allow us to also account for several further

¹⁵ 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>).

¹⁶ 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 (III_{nm}), manual skilled (III_m), semi-skilled (IV) and unskilled (V)). The Index of Multiple Deprivation (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.

measures of mother's health and behaviour, which may be correlated with both child height and the outcome of interest. We use two binary variables which measure whether the mother smoked or drank alcohol in the first three months of pregnancy; an ordered indicator for the intensity of mother's breastfeeding (never, <1 month, 1-3 months and 3+ months); mother's age at birth (20-24, 25-29, 30-34, 35+); mother's 'locus of control', a psychological concept that describes whether individuals attribute successes and failures to internal or external causes (those with an external locus of control attribute success and failure to chance); two further measures of maternal mental health; and finally several measures of parental involvement or interest in the child's development.¹⁷ We also use these covariates in a test of Assumption 1 above, examining whether they differ for the different genotypes.

4.4. Descriptive Statistics

Table 1 presents mean height (at age 8) for each of the SNPs, distinguishing between children who are homozygous for the height-increasing allele, heterozygous and homozygous for the height non-increasing allele. These show that each of the individual SNPs explain little of the variation in child height. This would imply that the first stage regressions have low explanatory power. For this reason (and as Weedon et al. (2008) and Lettre et al. (2009)), we create a count of the total number of height-increasing alleles carried by each child and use this as the instrument for child height. With nine SNPs, this variable has a possible range from 0 (homozygous for the non-height-increasing allele at all variants) to 18 (homozygous for the height-increasing allele at all variants). The left panel of Figure 2 presents a histogram of the number of 'tall' alleles carried by each child, showing a bell-shaped distribution. The linear prediction of child height, obtained from a regression on the number of 'tall' alleles, is presented by the straight line. On average, each 'tall' allele increases the child's height at age 8 by 0.043 standard deviations (approximately 0.25 cm). There is, however, a considerable amount of unexplained variation in height ($R^2 = 0.8\%$), as shown in the right panel of Figure 1, where the linear prediction is presented by the same straight line.

Columns 1 and 2 in Table 2 present the descriptive statistics (mean, standard deviation) of the variables used in the analyses. This shows an average height at age 8 of 132.2 cm and of 163.3 cm at

¹⁷ Maternal mental health is measured by 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 mental health, measuring general anxiety, depression and somaticism. Higher scores mean the mother is more affected. The mother's 'teaching score' 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. Finally, we also account for the extent to which parents engage in active (outdoor) activities with their children, such as going to the park or playground and going swimming.

age 13. In the analysis, we use standardised heights. Columns 3 to 5 show the raw association between this measure, the covariates and the number of ‘tall’ alleles, obtained from a regression of standardised height or each covariate on the number of height-increasing alleles. The top two rows of these columns present the relationship between child height and the instrument, showing a strong relationship for height at both ages. On average, each ‘tall’ allele is associated with a 0.043-0.047 standard deviation increase in child height (note, as mentioned above, that height is distributed with mean 100, standard deviation 10). The rest of columns (3) – (5) show no clear patterns or (with three exceptions) statistically significant associations in the relationship between the contextual variables and the number of height-increasing alleles. Using a two-sided binomial probability test at the 5% level, a comparison of the *observed* versus *expected* number of significant correlations suggests that the genetic variants show no greater association with the child and family background characteristics than what would be expected by chance ($p = 0.15$), providing support for Assumption 1.¹⁸

4.5 IV Falsification Check

Another way to examine the robustness of our IV approach and the validity of our instruments is by undertaking a ‘falsification check’. If we examine the effect of height on body weight, controlling for other covariates, we expect (a) to find strong positive effects in both OLS and IV, as the two are highly (positively) correlated particularly in children who are still growing (e.g. see any children’s growth charts), and (b) for the two estimates to give similar coefficients. There would be a slight downward bias of the OLS estimates in some case; for example, a healthy (unobserved) diet might positively affect height and negatively affect weight. As shown by Tanner (1978) and discussed above however, even with severe (prenatal or postnatal) malnutrition, children attain similar heights as their siblings or peers. Hence, even if any bias exists, we do not expect this to be large. A substantially different or null IV finding would therefore cast doubt on our IV strategy.

Table 3 shows strong positive estimates of body weight on height at different ages in both the OLS and IV. A one standard deviation increase in height is associated with a 0.52 – 0.69 standard deviation increase in weight in the OLS, and a 0.26 – 0.98 standard deviation increase in weight in the IV. The point estimates are similar in both models, though the standard errors are much larger in the IV as expected. Using the Durbin-Wu-Hausman test, we can statistically distinguish the OLS from IV in

¹⁸ To shed more light on whether the variants are likely to be related to other background characteristics, we also examine the relationship between the genetic variants and a wide set of further variables (64 additional pairwise comparisons) that are not included in our analysis (such as whether the child had sleeping difficulties, the child’s ‘locus of control’, whether the mother had a caesarean section, mother’s self-esteem, anxiety, depression, whether the family owns their own home, whether they have financial difficulties, etc.). The findings (available from the authors upon request) also suggest the genetic variants are unrelated to these other variables (using a two-sided binomial probability test, $p = 0.77$ at the 5% level; $p = 0.83$ at the 10% level).

only one case (11 year-old girls); the majority of the IV estimates are indistinguishable from those estimated by OLS.

This suggests that our instruments perform well and that the IV approach correctly identifies the causal effect of height on body weight. Although this does not guarantee that our IV approach also correctly identifies the causal effect on the other outcomes of interest, it does provide support for our argument that both the approach and the instruments are valid to obtain causal estimates of the effects of stature.

5. Results

5.1. OLS Results

We begin by examining the OLS association between height, cognitive skills and mental health. Columns (1) and (2) of Table 4 show a positive association between height, test scores and IQ that halves when controlling for the background characteristics. However, the actual magnitude of the association is small: controlling for all covariates (the 'adjusted' results), a one standard deviation increase in height is associated, for example, with a 0.056 standard deviation increase in girls' IQ. Comparing this to the effect of within-school-year age on IQ in our data, this corresponds to a difference in test scores between children born approximately one month apart.

Columns (3) to (5) examine the relationship between height, the two measures of self-esteem and symptoms of depression. This shows that height is related to increases in self-esteem and depression scores, though the estimates decrease and become indistinguishable from the null when controlling for the covariates (the positive association with depression symptoms within girls is the one exception).

Table 5 presents both the unadjusted (raw) and adjusted (controlling for all covariates) associations between height and behavioural problems. The latter shows that height is unrelated to hyperactivity and conduct problems, but there is a negative relationship with emotional problems. The effects are again small: a one standard deviation increase in height is associated with 0.06-0.07 standard deviations decrease in emotional problems. The results also show a small negative association between height and peer problems for girls.

5.2. IV results

Table 6 presents the IV results for cognitive skills and mental health. The estimates are obtained from a pooled (by gender) first stage regression of height (e.g. at age 8) on the number of height-increasing alleles carried by each child. Our instrument predicts height well, with a first stage F-statistic between 34 and 42, satisfying Assumption 2.¹⁹ To obtain gender-specific second stage results, we regress the outcome on the predicted height interacted with the child's gender, bootstrapping the standard errors.

Column 1, Table 6, shows the IV estimates for KS3. These are larger than the OLS but also have large standard errors. We cannot reject the null of no effect, nor can we reject the Durbin-Wu-Hausman (DWH) test, which is a test of the hypothesis that the IV estimates are the same as those in the OLS. Column 2 shows a positive effect of instrumented height on IQ for both boys and girls and the DWH test rejects the null of exogenous height at the 10% level. The IV estimate is larger than the OLS, suggesting that the latter underestimates the true effect: a one standard deviation increase in instrumented height increases IQ by 0.3 standard deviations (we discuss possible reasons for this difference below).

The IV estimates suggest, in contrast to the OLS estimates, that height is negatively related to the mental health outcomes. Columns 3 to 5 of Table 6 show that for self-esteem, global self worth and depression symptoms, the large standard errors mean we cannot reject the null of no effect. But all three sets of IV coefficients relate increasing height to worse outcomes.

Table 7 presents the IV regression results for behavioural problems. In contrast to the OLS results in Table 5, the IV estimates in Column 1 of Table 7 show height to be a predictor of hyperactivity in girls, with the DWH test rejecting the exogeneity assumption of height. A one standard deviation increase in instrumented height increases the hyperactivity score by 0.34 standard deviations. The estimate for boys is also positive, though smaller and not statistically distinguishable from the null. Similarly, height appears to be a positive predictor of boys' emotional and peer problems (columns 2 and 4) and the estimated effect, although not statistically significant, is of a similar magnitude for girls' emotional problems. Column 3 shows there is no effect of height on conduct problems for either gender.

¹⁹ 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 estimate the first stage regression, interacting the genetic variants with indicators for various subgroups and test whether the instrument coefficient is 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. With one interaction (breastfeeding for 11+ girls, but not for boys) showing $p=0.057$, and all other p -values between 0.138 and 0.977, the results (available from the authors) show no more significant differences than what would be expected by chance, suggesting that gene-environment interactions do not play an important role for the genetic variants used here.

5.3 Non-linearities

As discussed in section 2, the existing literature has found both tallness and shortness to have negative psychological effects in children. The estimates discussed above only examine differences in the outcome of interest at the mean, but the relationship between height and the outcomes may differ at different points in the distribution. We therefore investigate different cut-points and examine the effects of being below the 25th and above the 75th percentile of the age- and gender specific height distribution. The results (available upon request) confirm our main findings. OLS estimates show that shorter children have lower IQ and do worse in school tests, and vice versa for taller children. The IV results also show that shorter children have lower IQ. But, as above, the IV estimates show no evidence of a relationship between height and scholastic competence, self-worth or depression. The IV effects of being tall or short on the child's behavioural problems also show similar patterns to those above: relatively short girls are less hyperactive and have fewer emotional problems, and vice versa for tall girls. For boys, shortness decreases and tallness increases emotional and peer problems.

6. Conclusion and discussion

This paper is the first to exploit genetic variation in height to examine the causal effects of height on human capital accumulation. OLS results show that taller children perform better in terms of cognitive performance and are less likely to have emotional and peer problems (girls), though tall girls are more likely to show symptoms of depression. Using genetic variation in height in an IV specification, we attempt to deal with the problems of endogeneity. The IV findings are similar to the OLS for cognitive performance; we find a positive effect of height on IQ. However, we find a negative relationship with behaviour. We show that the OLS results are downwardly biased and that height *increases* rather than *decreases* behavioural problems. Taller children are more hyperactive and are more likely to have emotional and peer problems.

In many of our results, the IV estimates indicate that OLS is biased downwards. One possible explanation for the difference between IV and OLS could be a genetic one. For example, (one of) our SNPs could be pleiotropic or in LD with another variant that directly affects IQ or cognition. But from the evidence discussed in footnote 11 and from the fact that we use only nine SNPs out of possibly hundreds or thousands SNPs coding for height, we make the assumption that this is not the case. More generally, the literature does not give any reason to expect violation of the IV assumption

based on the known function of these genes. In addition, there is no reason to expect violation given our tests of associations with known confounders and our falsification check. However, as we cannot directly test this, this has to remain an assumption.²⁰

One possible explanation for our IV findings that indicate that being taller *increases* rather than *decreases* behavioural problems could be the differential treatment of children of different stature. A ‘size-appropriate’ rather than ‘age-appropriate’ treatment of tall children may trigger behavioural problems. Expectations and reactions to ‘tall-for-age’ children’s (what may seem childish) behaviour can in turn affect children’s development. As factors such as socio-economic position are positively related to height and negatively related to behavioural problems, the OLS estimates will be downward biased if these factors are insufficiently controlled for. Though possible, these are speculations as we currently have no further evidence to confirm these. However, the finding of increased behavioural problems is consistent with the psychological literature that has shown a positive relationship between height and children’s behavioural problems, though this literature has mainly examined outcomes such as aggression and violence (Raine et al., 1998; Farrington, 1989) rather than those we examine here.

Finally, the IV effects we find for behaviour and IQ are large: a one standard deviation increase in height raises these scores by about 0.3 standard deviations. This is a similar magnitude to the lower bound of our estimate of the causal effect of height on weight, suggesting that these effects are not trivial. In fact, comparison of these effects with those of other child characteristics shows they are large. For example, a 0.34 standard deviation difference in IQ is comparable to the difference in this score for children born approximately 6 months apart within the same school year. Likewise, the difference between girls’ and boys’ raw hyperactivity scores is approximately 0.37 standard deviations which is similar to the estimated effect of one standard deviation increase in height on hyperactivity.

In conclusion, our findings suggest that height is an important factor in children’s human capital accumulation in both childhood and adolescence, likely as a result of the social reactions that are triggered by variations in height. We show that being tall may not only confer advantage but also disadvantage; our examination of behavioural problems contrasts with the more positive view of height that emerges from the existing empirical literature on height and children’s cognitive performance.

²⁰ As discussed in Davey Smith (in press), genetic confounding can be tested indirectly by obtaining many IV estimates that use different independent (combinations of) variants. These estimates will not plausibly be influenced by any common pleiotropy or LD-induced confounding. Therefore, if they display consistency, it provides evidence against genetic confounding. Using this approach, we specified each variant separately as the instrumental variable. This greatly reduces the power in the first stage, and generally more than doubles the standard errors in the second stage, but it broadly supports the findings reported above (results available from the authors upon request).

References

- Angeles, I. et al. (1993). "Decreased Rate of Stunting among Anemic Indonesian Preschool Children through Iron Supplementation." *Amer J of Clin Nutr* 58(3): 339-42.
- Angold, A. et al. (1995). "Development of a Short Questionnaire for Use in Epidemiological Studies of Depression in Children and Adolescents." *International Journal of Methods in Psychiatric Research*, 5(4): 237-249.
- Angrist, J., and J-S Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion* (Princeton: Princeton University Press).
- Angrist, J., K. Graddy, and G. Imbens. (2000). "The Interpretation of Instrumental Variables Estimators in Simultaneous Equation Models with an Application to the Demand for Fish," *Review of Economic Studies* 67, 499-527
- Axelrad D., et al. (2007). "Dose-response Relationship of Prenatal Mercury Exposure and IQ: An Integrative Analysis of Epidemiologic Data." *Environmental Health Perspectives* 115:609–15.
- Cohen J, et al. (2005). "A Quantitative Analysis of Prenatal Methyl Mercury Exposure and Cognitive Development". *American Journal of Preventive Medicine* 29:353–65.
- Goodman, R. (1997). "The Strengths and Difficulties Questionnaire: A Research Note." *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 38(5), 581-586.
- Behrman, J. et al. (1994). "Endowments and the Allocation of Schooling in the Family and in the Marriage Market: The Twins Experiment." *Journal of Political Economy* 102(6):1131-74
- Bhatti P, et al. (2005). "Genetic Variation and Willingness to Participate in Epidemiologic Research: Data from Three Studies." *Cancer Epidemiology Biomarkers and Prevention* 14:2449-53.
- Binder, G. et al. (1997). "Outcome in Tall Stature. Final Height and Psychological Aspects in 220 Patients with and without Treatment." *European Journal of Paediatrics* 156:905-10.
- Bound, J. and G. Solon (1999). "Double Trouble: On the Value of Twins-Based Estimation of the Returns to Schooling." *Economics of Education Review* 18:169-82.
- Case, A. and C. Paxson (2008). "Stature and status: Height, Ability and Labor Market Outcomes" *Journal of Political Economy* 116(3): 499-532.
- Case, A. and C. Paxson (2010). "Causes and Consequences of Early Life Health" *NBER Working Paper* 15637.
- Clifford, M. and E. Walster (1973). "The Effect of Physical Attractiveness on Teacher Expectations." *Sociology of Education* 46(2): 248-58.
- Darley, J. and R. Fazio (1980). "Expectancy Confirmation Processes Arising in the Social Interaction Sequence." *American Psychologist* 35(10): 867-81.
- Davey Smith, G. and S. Ebrahim (2003). "'Mendelian Randomization': Can Genetic Epidemiology Contribute to Understanding Environmental Determinants of Disease?" *International Journal of Epidemiology* 32: 1-22.
- Davey Smith G. et al. (2008). "Clustered Environments and Randomized Genes: A Fundamental Distinction between Conventional and Genetic Epidemiology." *PLoS Medicine* 4: 1985-92.
- Davey Smith G. (2008). "Assessing Intrauterine Influences on Offspring Health Outcomes: Can Epidemiological Findings yield Robust Results?" *Basic and Clinical Pharmacology and Toxicology* 102:245-56.
- Davey Smith G. (*in press*). "Use of Genetic Markers and Gene-Diet Interactions for Interrogating Population-Level Causal Influences of Diet on Health." *Genes and Nutrition*
- Dipboye, R. et al. (1975). "Relative Importance of Applicant Sex, Attractiveness and Scholastic Standing in Evaluation of Job Applicant Resumes." *Journal of Applied Psychology* 60(1): 39-43.
- Elgen, I., et al. (2002). "Population Based, Controlled Study of Behavioural Problems and Psychiatric Disorders in Low Birthweight Children at 11 Years of Age." *Archives of Disease in Childhood: The Fetal and Neonatal Edition* 87(2):F128-32.
- Ericson, A. and B. Kallen (1998). "Very Low Birthweight Boys at Age 19." *Archives of Disease in*

- Childhood - Fetal and Neonatal Edition* 78(May): F171-4.
- Farrington, D. (1989). "Early Predictors of Adolescent Aggression and Adult Violence." *Violence and Victims* 4(2):79-100.
- Gilman, S., H. Gardener and S. Buka (2008). "Maternal Smoking during Pregnancy and Children's Cognitive and Physical Development: A Causal Risk Factor?" *American Journal of Epidemiology* 168(5): 522-31.
- Golding, J. et al. (2001). "ALSPAC - The Avon Longitudinal Study of Parents and Children: I. Study Methodology." *Pediatric and Perinatal Epidemiology* 15: 74-87.
- Griliches, Z. (1979). "Sibling Models and Data in Economics: Beginnings of a Survey." *Journal of Political Economy* 87(5):S37-S64.
- Gudbjartsson, D. et al. (2008). "Many Sequence Variants affecting Diversity of Adult Human Height." *Nature Genetics* 40(5): 609-15.
- Hamermesh, D. and J. Biddle (1994). "Beauty and the Labor Market." *American Economic Review* 84(5): 1174-94.
- Harter, S. (1985). "Self-perception profile for children". University of Denver.
- Hensley, W. (1993). "Height as a Measure of Success in Academe." *Psychology: A Journal of Human Behavior* 30: 40-6.
- Hinke Kessler Scholder, von. S. et al. (2010). "Genetic Markers as Instrumental Variables: An Application to Child Fat Mass and Academic Achievement." *CMPO Working Paper 10/229*
- Jones, M. and N. Bayley (1950). "Physical Maturing among Boys as Related to Behavior." *The Journal of Educational Psychology* 41(3):129-48.
- Judge, T. and D. Cable (2004). "The Effect of Physical Height on Workplace Success and Income: Preliminary Test of a Theoretical Model." *Journal of Applied Psychology* 89(3): 428-41.
- Kafouri, S. et al. (2009). "Maternal cigarette smoking during pregnancy and cognitive performance in adolescence." *International Journal of Epidemiology* 38(1):158-72.
- Kivimäki, M. et al. (2008). "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): 2552-60.
- Lawlor, D. et al. (2008). "Mendelian Randomization: Using Genes as Instruments for Making Causal Inferences in Epidemiology." *Statistics in Medicine* 27: 1133-63.
- Lettre, G. et al. (2008). "Identification of Ten Loci associated with Height highlights New Biological Pathways in Human Growth." *Nature Genetics* 40(5): 584-91.
- Ligon, A. et al. (2005). "Constitutional Rearrangement of the Architectural Factor HMGA2: A Novel Human Phenotype including Overgrowth and Lipomas." *American Journal of Human Genetics*, 76(2):340-8.
- Lozoff, B., et al. (2006). "Long-Lasting Neural and Behavioral Effects of Iron Deficiency in Infancy." *Nutrition Reviews* 64(1):34-43.
- Macgregor, S. et al. (2006). "Bias, Precision and Heritability of Self-Reported and Clinically Measured Height in Australian Twins." *Human Genetics*, 120(4):571-80.
- Macintyre, S. and P. West (1991). "Social, Developmental and Health Correlates of 'Attractiveness' in Adolescence." *Sociology of Health and Illness* 13(2): 149-67.
- Magnusson, P., F. Rasmussen and U. Gyllensten (2006). "Height at Age 18 is a Strong Predictor of Attained Education Later in Life: Cohort Study of over 950000 Swedish Men." *International Journal of Epidemiology* 35: 658-63.
- McEvoy, B. and P. Visscher (2009). "Genetics of Human Height." *Economics and Human Biology*, 7:294-306.
- Mills, J. et al. (1984). "Maternal Alcohol Consumption and Birth Weight. How Much Drinking During Pregnancy is Safe?" *JAMA* 252(14): 1975-9.
- Molfese, V. et al. (1997). "Prediction of the Intelligence Test Scores of 3- to 8-year old Children by Home Environment, Socioeconomic Status, and Biomedical Risks." *Merrill-Palmer Quarterly*, 43, 219-34.
- Morison, P. and A. Masten (1991). "Peer Reputation in Middle Childhood as a Predictor of Adaptation

- in Adolescence: A Seven-year Follow-up." *Child Development* 62:991-1007.
- Nilsson, P. (2008). "Does a Pint a Day Affect Your Child's Pay? The Effect of Prenatal Alcohol Exposure on Adult Outcomes." *IFAU Working Paper No. 4*.
- Oken, E. and D. Bellinger (2008). "Fish Consumption, Methylmercury and Child Neurodevelopment." *Current Opinion in Pediatrics* 20(2):178–83
- Olds, D. et al. (1994). "Intellectual Impairment in Children of Women who Smoke Cigarettes During Pregnancy." *Pediatrics* 93(2): 221-7.
- Parker, J. and S. Asher (1987). "Peer Relations and Later Personal Adjustment: Are Low-Accepted Children at Risk?" *Psychological Bulletin* 102(3):458-9.
- Persico, N., A. Postlewaite and D. Silverman (2004). "The Effect of Adolescent Experience on Labour Market Outcomes: The Case of Height." *NBER Working paper 10522*.
- Pyett, P. et al. (2005). "Using Hormone Treatment to Reduce the Adult Height of Tall Girls: Are Women Satisfied with the Decision in Later Years?" *Social Science and Medicine* 61:1629-39.
- Raine, A. et al. (1998). "Fearlessness, Stimulation-Seeking, and Large Body Size at Age 3 as Early Predispositions to Childhood Aggression at Age 11 Years." *Archives of General Psychiatry* 55:745-51.
- Richards, M. et al. (2002). "Birthweight, Postnatal Growth and Cognitive Function in a National UK Birth Cohort." *International Epidemiological Association* 31: 342-8.
- Rotnem, D. et al. (1977). "Personality Development in Children with Growth Hormone Deficiency." *Journal of the American Academy of Child and Adolescent Psychiatry* 16(3):412-26.
- Royston, P. (2004). "Multiple Imputation of Missing Values." *Stata Journal* 4(3):227-41.
- Russell, M. (1991) "Clinical Implications of Recent Research on the Fetal Alcohol Syndrome.", *Bulletin of the New York Academy of Medicine*, 67: 207-222.
- Sanna, S. et al. (2008). "Common Variants in the GDF5-UQCC Region are Associated with Variation in Human Height." *Nature Genetics*, 40(2):198-203.
- Sandberg, D. and L. Voss (2002). "The Psychosocial Consequences of Short Stature: A Review of the Evidence." *Best Practice and Research Clinical Endocrinology and Metabolism* 16(3):449-63.
- Sandberg, D. et al. (2004). "Height and Social Adjustment: Are Extremes a Cause for Concern and Action?" *Pediatrics* 114(3):744-50.
- Silventoinen, K. et al. (2000). "Genetic and Environmental Contributions to the Association between Body Height and Educational Attainment: A Study of Adult Finnish Twins." *Behavior Genetics* 30(6): 477-485.
- Silventoinen, K. et al. (2003). "Heritability of Adult Body Height: A Comparative Study of Twin Cohorts in Eight Countries." *Twin Research* 6(5): 399-408.
- Skuse, D. et al. (1994). "Psychosocial Assessment of Children with Short Stature: A Preliminary Report." *Acta Paediatrica Suppl.* 83(s406):11-6.
- Sundet, J. et al. (2005). "Resolving the Genetic and Environmental Sources of the Correlation between Height and Intelligence: A Study of Nearly 2600 Norwegian Male Twin Pairs." *Twin Research and Human Genetics* 8(4): 307-11.
- Tanner, J. (1978). *Foetus into Man: Physical Growth from Conception to Maturity*. London, Open Books Publishing Ltd.
- Tanner, J. (1982). "The Potential of Auxological Data for Monitoring Economic and Social Well-Being." *Social Science History* 6(4): 571-81.
- Underwood, L. (1991). "The Societal Cost of Being Short: Societal Perceptions and Biases." *Acta Paediatrica Scandinavica* 377, 3-8.
- Voss, L. and J. Mulligan (2000). "Bullying in School: Are Short Pupils at Risk? Questionnaire Study in a Cohort." *BMJ* 320: 612-3.
- Wechsler D, et al. (1992). "WISC-III^{UK} Wechsler Intelligence Scale for Children – Third Edition UK Manual." Sidcup, UK: The Psychological Corporation.
- Weedon, M. et al. (2007). "A Common Variant of *HMGA2* is Associated with Adult and Childhood Height in the General Population." *Nature Genetics* 39(10): 1245-50.

- Weedon, M. et al. (2008). "Genome-Wide Association Analysis Identifies 20 Loci that Influence Adult Height." *Nature Genetics* 40: 575-83.
- Wentzel, K. (2009). "Peers and Academic Functioning at School." In: Rubin, K. et al. (ed.) *Handbook of Peer Interactions, Relationships and Groups*. New York: The Guildford Press 531-47.
- West, P. (1991). "Rethinking the Health Selection Explanation for Health Inequalities." *Social Science and Medicine* 32(4): 373-84.
- Yang, J. et al. (2008). "Birth Weight and Cognitive Ability in Childhood among Siblings and Nonsiblings." *Pediatrics* 122(2):e350-e358.
- Yang, J. et al. (2010). "Common SNPs explain a Large Proportion of the Heritability for Human Height." *Nature Genetics* 42:565-71.

Figures and Tables

Figure 1: A directed acyclic graph (DAG) showing that the effect of height H_i on the outcome of interest C_i can run via different pathways P . The unobservables e_i may be related to both H_i and C_i . The instrument Z_i only affects C_i via its effect on H_i .

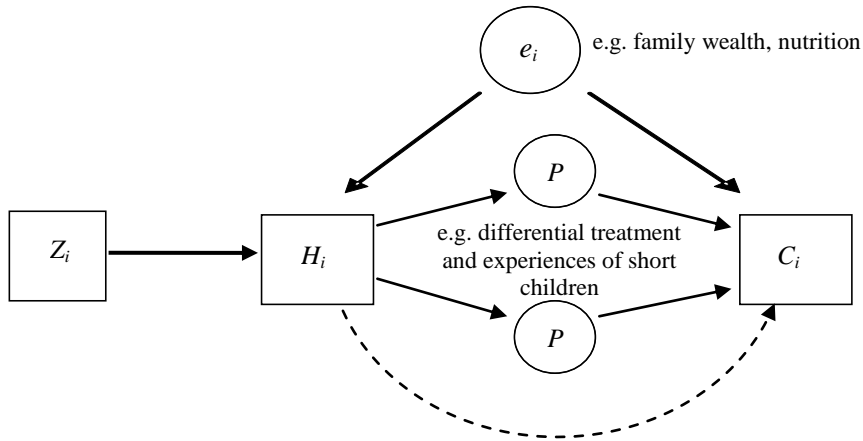


Figure 2: Histogram of children's height at age 8 by the number of height-increasing alleles

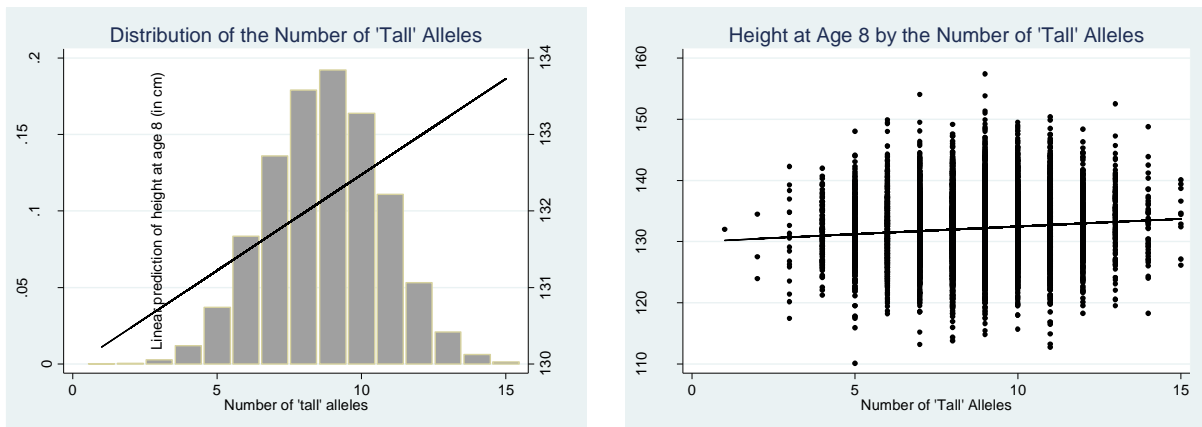


Table 1: Mean and standard deviation of height at age 8 (in cm) for each SNP.

SNP	rs number	'tall' allele	Homozygous for 'non-tall' allele		Heterozygous		Homozygous for 'tall' allele	
			Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
<i>HMGA2</i>	rs1042725	<u>C</u> /T	132.0	(5.47)	132.0	(5.59)	132.7	(5.71)
<i>ZBTB38</i>	rs6440003	<u>A</u> /G	131.8	(5.53)	132.3	(5.68)	132.4	(5.47)
<i>GDF5</i>	rs6060373	<u>C</u> /T	132.2	(5.58)	132.2	(5.59)	132.2	(5.71)
<i>LOC387103</i>	rs4549631	<u>C</u> /T	131.6	(5.45)	132.5	(5.68)	132.1	(5.54)
<i>EFEMP1</i>	rs3791675	<u>A</u> /G	131.0	(5.75)	131.9	(5.52)	132.4	(5.61)
<i>SCMH1</i>	rs6686842	<u>A</u> /G	131.9	(5.51)	132.2	(5.62)	132.4	(5.68)
<i>ADAMTSL3</i>	rs10906982	<u>A</u> /T	131.8	(5.53)	132.1	(5.64)	132.5	(5.57)
<i>DYM</i>	rs8099594	<u>T</u> /C	131.9	(5.72)	132.2	(5.51)	132.2	(5.65)
<i>C6orf106</i>	rs2814993	<u>T</u> /C	132.2	(5.70)	132.1	(5.34)	132.5	(5.19)

Note: The height-increasing allele is bold and underlined

Table 2: Descriptive statistics of control variables: Columns 1-2 show their mean and standard deviation. Columns 3-5 present the coefficients, standard error and p-value of the variables shown in the first column regressed on the instrument (a count of the number of height-increasing alleles).

	(1) Mean	(2) Std. dev.	(3) Coeff.	(4) Std. err.	(5) p-value
Height (columns 1-2: in cm; columns 3-5: standardised)					
Age 8	132.2	(5.6)	0.430	0.081	<0.000
Age 13	163.3	(7.6)	0.474	0.089	<0.000
Control variables					
Age in months at Focus@8 clinic	103.1	(2.21)	0.018	0.017	0.284
Age in months at Teen Focus 2 clinic	166.0	(2.12)	0.011	0.015	0.466
Age in months at KS3 exam	169.6	(3.76)	-0.033	0.025	0.194
Birth weight (g)	3422	(549)	-1.505	3.699	0.684
Younger siblings under 18 in the household	0.51	(0.65)	-0.003	0.004	0.543
Older siblings under 18 in the household	0.74	(0.74)	0.000	0.005	0.972
Ln(income)	5.32	(0.45)	0.000	0.003	0.995
Father's education	2.43	(1.01)	0.000	0.007	0.986
Mother's education	2.36	(0.88)	0.003	0.006	0.615
Mother's mother's education	1.73	(0.75)	-0.001	0.005	0.829
Mother's father's education	1.84	(0.79)	-0.003	0.005	0.610
Child is not raised by natural father	0.06	(0.23)	0.001	0.002	0.628
Father's social class at the child's birth	3.04	(1.28)	-0.001	0.009	0.894
Mother is employed part-time at 21 months	0.40	(0.49)	0.008	0.003	0.020
Mother is employed full-time at 21 months	0.10	(0.30)	-0.000	0.002	0.906
Father is employed at 21 months	0.92	(0.28)	0.004	0.002	0.038
Index of Multiple Deprivation (IMD)	19.50	(13.95)	-0.014	0.100	0.891
Mother: alcohol in month 1-3 of pregnancy	0.57	(0.50)	0.004	0.003	0.266
Mother: smoked in month 1-3 of pregnancy	0.18	(0.39)	-0.000	0.003	0.917
Breastfeeding	1.88	(1.20)	0.011	0.008	0.207
Mother's age	3.38	(0.91)	0.002	0.006	0.711
Mother's 'locus of control'	98.88	(9.44)	-0.020	0.065	0.764
Mother's EPDS	6.46	(4.54)	0.025	0.031	0.412
Mother's CCEI	12.88	(7.19)	0.038	0.049	0.443
Teaching score	7.02	(0.93)	0.001	0.007	0.916
Activities (indoor) score	0.69	(0.20)	0.003	0.001	0.032
Activities (outdoor) score	27.89	(4.61)	-0.008	0.032	0.810

Note: Rather than height in cm, the analysis uses standardised heights (with mean 100, standard deviation 10).

Table 3: Falsification check: The effects of contemporaneous height on body weight, OLS and IV

	Weight (measured at the same age as height)			
	OLS		IV	
	Boys	Girls	Boys	Girls
Height, age 3	0.645*** (0.016)	0.614*** (0.016)	0.638*** (0.122)	0.576*** (0.124)
p-value DWH test			0.954	0.757
No. of observations	5248		5248	
Height, age 5	0.571*** (0.027)	0.569*** (0.031)	0.981*** (0.305)	0.822*** (0.271)
p-value DWH test			0.177	0.347
No. of observations	3359		3359	
Height, age 8	0.673*** (0.018)	0.689*** (0.020)	0.543 *** (0.122)	0.467*** (0.142)
p-value DWH test			0.281	0.114
No. of observations	4335		4335	
Height, age 11	0.648*** (0.018)	0.637*** (0.017)	0.469*** (0.119)	0.260** (0.114)
p-value DWH test			0.128	0.001
No. of observations	4442		4442	
Height, age 13	0.600*** (0.017)	0.555*** (0.023)	0.637*** (0.135)	0.333** (0.146)
p-value DWH test			0.782	0.124
No. of observations	3865		3865	
Height, age 15	0.609*** (0.024)	0.521*** (0.029)	0.792*** (0.180)	0.568*** (0.160)
p-value DWH test			0.305	0.765
No. of observations	3383		3383	

Notes: The estimates come from regressions of body weight on contemporaneous height interacted with gender; controls include: birth weight, age in months, number of older and younger siblings, log equivalised family income and its square, mother's -, father's -, and mother's parents' educational level, raised by natural father, social class, maternal age at birth, parents' employment status, IMD at birth, mother's smoking and drinking during pregnancy, breastfeeding, mother's 'locus of control' and mental health (EPDS and CCEI), parental involvement or interest in the child's development, and parents' engagement in active (outdoor) activities with their child; DWH test = Durbin-Wu-Hausman test ($H_0: \beta_1$ is exogenous); * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: OLS – The unadjusted and adjusted effects of contemporaneous height (ages 8 and 13) on cognitive skills and mental health

	(1) Key Stage 3, Age 14		(2) IQ test score, Age 8		(3) Scholastic self-esteem, Age 8		(4) Global self-worth, Age 8		(5) Depression, Age 13	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
<u>Unadjusted</u>										
Height	0.091*** (0.016)	0.106*** (0.017)	0.113*** (0.015)	0.111*** (0.015)	0.029* (0.015)	0.030* (0.016)	0.019 (0.016)	0.027* (0.016)	0.011 (0.016)	0.048*** (0.017)
<u>Adjusted for all covariates</u>										
Height	0.039** (0.017)	0.084*** (0.022)	0.063*** (0.020)	0.056*** (0.018)	0.008 (0.023)	0.020 (0.022)	0.012 (0.023)	0.020 (0.022)	-0.004 (0.019)	0.054* (0.030)
No. of obs.	3149		4522		4270		4270		3828	

Notes: The estimates come from regressions of the outcome on contemporaneous height interacted with gender; controls are listed in the note to Table 3; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: OLS – The unadjusted and adjusted effects of contemporaneous height on behaviour at age 13

	(1) Hyperactivity		(2) Emotional problems		(3) Conduct problems		(4) Peer problems	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
<u>Unadjusted</u>								
Height (age 13)	0.003 (0.018)	-0.034* (0.018)	-0.076*** (0.017)	-0.058*** (0.018)	-0.004 (0.017)	-0.012 (0.017)	0.010 (0.018)	-0.029 (0.019)
<u>Adjusted for all covariates</u>								
Height (age 13)	-0.002 (0.023)	0.022 (0.027)	-0.060*** (0.020)	-0.070** (0.032)	0.001 (0.021)	0.009 (0.028)	0.017 (0.024)	-0.057* (0.028)
No. of obs.	3315		3314		3314		3312	

Notes: The estimates come from regressions of the outcome on contemporaneous height interacted with gender; Controls are listed in the note to Table 3; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: IV – The (adjusted) effects of contemporaneous height (ages 8 and 13) on cognitive skills and mental health

	(1) Key Stage 3, Age 14		(2) IQ test score, Age 8		(3) Scholastic self-esteem, Age 8		(4) Global self-worth, Age 8		(5) Depression, Age 13	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Height at age 8	0.061 (0.130)	0.153 (0.127)	0.352** (0.163)	0.344** (0.174)	-0.063 (0.184)	-0.102 (0.203)	-0.088 (0.176)	-0.103 (0.167)	0.103 (0.183)	0.249 (0.175)
p-value DWH test	0.885	0.549	0.074	0.096	0.697	0.545	0.567	0.457	0.557	0.258
First stage <i>F</i> -statistic	34.2		40.0		38.5		38.5		42.0	

Notes: Estimates are obtained from a first stage regression of height (e.g. at age 8) on the number of height-increasing alleles carried by each child. The second stage regresses the outcome on the predicted height interacted with a variable indicating the child's gender. Standard errors are obtained using bootstrapping; All controls included, these are listed in the note to Table 3; DWH test = Durbin-Wu-Hausman test ($H_0: \beta_1$ is exogenous); * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: IV – The (adjusted) effects of contemporaneous height on behaviour at age 13

	(1) Hyperactivity		(2) Emotional problems		(3) Conduct problems		(4) Peer problems	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Height at age 13	0.168 (0.141)	0.340** (0.156)	0.322* (0.178)	0.256 (0.203)	-0.231 (0.192)	-0.010 (0.198)	0.312* (0.172)	-0.010 (0.160)
p-value DWH test	0.222	0.038	0.031	0.104	0.224	0.923	0.083	0.765
<i>F</i> -statistic	38.4		38.9		39.0		38.7	

Notes: Estimates are obtained from a first stage regression of height (e.g. at age 8) on the number of height-increasing alleles carried by each child. The second stage regresses the outcome on the predicted height interacted with a variable indicating the child's gender. Standard errors are obtained using bootstrapping; All controls included; these are listed in the notes to Table 3. DWH test = Durbin-Wu-Hausman test ($H_0: \beta_1$ is exogenous); * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Appendix A: A Brief Introduction to Genetics

Each cell in the human body contains a nucleus in which most DNA (99.9995%) is kept. 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 *HMGA2* SNP (one of the genetic variants used here): they can be homozygous for the common allele (TT), heterozygous (CT), and homozygous for the rare allele of *HMGA2* (CC). The visible or measurable effect of a particular genotype is called the phenotype.

The phenotype we examine in this paper is child height. Studies that examine the heritability of human height generally report large proportions of the variance that are due to genetics: up to 0.9 (Silventoinen et al., 2003; Macgregor *et al.*, 2006).²¹ 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 height, though all with small effects: so-called 'polygenes'. Together, these variants may have a large phenotypic effect.

²¹ 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 the characteristic in monozygotic (MZ) twins with intra-pair correlation in dizygotic (DZ) twins. The implicit assumption is that the effects of shared environmental factors are similar for MZ and DZ twins. The heritability of a characteristic is calculated as twice the difference between MZ and DZ intra-pair correlations ($h^2 = 2*(r_{MZ} - r_{DZ})$).