

**This is an electronic reprint of the original article.
This reprint *may differ* from the original in pagination and typographic detail.**

Author(s): Böckerman, Petri; Viinikainen, Jutta; Vainiomäki, Jari; Hintsanen, Mirka; Pitkänen, Niina; Lehtimäki, Terho; Pehkonen, Jaakko; Rovio, Suvi; Raitakari, Olli

Title: Stature and long-term labor market outcomes : Evidence using Mendelian randomization

Year: 2017

Version:

Please cite the original version:

Böckerman, P., Viinikainen, J., Vainiomäki, J., Hintsanen, M., Pitkänen, N., Lehtimäki, T., Pehkonen, J., Rovio, S., & Raitakari, O. (2017). Stature and long-term labor market outcomes : Evidence using Mendelian randomization. *Economics and Human Biology*, 24, 18-29. <https://doi.org/10.1016/j.ehb.2016.10.009>

All material supplied via JYX is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Accepted Manuscript

Title: Stature and long-term labor market outcomes: Evidence using Mendelian randomization

Author: Petri Böckerman Jutta Viinikainen Jari Vainiomäki
Mirka Hintsanen Niina Pitkänen Terho Lehtimäki Jaakko
Pehkonen Suvi Rovio Olli Raitakari



PII: S1570-677X(16)30046-6
DOI: <http://dx.doi.org/doi:10.1016/j.ehb.2016.10.009>
Reference: EHB 609

To appear in: *Economics and Human Biology*

Received date: 3-6-2016
Revised date: 21-10-2016
Accepted date: 27-10-2016

Please cite this article as: Böckerman, Petri, Viinikainen, Jutta, Vainiomäki, Jari, Hintsanen, Mirka, Pitkänen, Niina, Lehtimäki, Terho, Pehkonen, Jaakko, Rovio, Suvi, Raitakari, Olli, Stature and long-term labor market outcomes: Evidence using Mendelian randomization. *Economics and Human Biology* <http://dx.doi.org/10.1016/j.ehb.2016.10.009>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**Stature and long-term labor market outcomes:
Evidence using Mendelian randomization**

Petri Böckerman^{a,*} Jutta Viinikainen^{b,**} Jari Vainiomäki^{***} Mirka Hintsanen,^a Niina Pitkänen^b, Terho Lehtimäki^c, Jaakko Pehkonen^d, Suvi Rovio^e, Olli Raitakari^f

^aTurku School of Economics, Labour Institute for Economic Research, Helsinki, Finland and IZA. Phone: +358-9-2535 7330. Fax: +358-9-2535 7332. E-mail: petri.boeckerman@labour.fi

^bJyväskylä University School of Business and Economics, Jyväskylä, Finland.

^cSchool of Management, University of Tampere, Finland.

^dUnit of Psychology, University of Oulu, Oulu, Finland.

^eResearch Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Finland.

^fDepartment of Clinical Chemistry, Fimlab Laboratories and School of Medicine, University of Tampere, Finland.

^gJyväskylä University School of Business and Economics, Jyväskylä, Finland.

^hResearch Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland.

ⁱResearch Centre of Applied and Preventive Cardiovascular Medicine, University of Turku and Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland.

*Corresponding author.

Highlights

- Taller people have higher earnings.
- We use genetic instruments for height.
- The OLS models show that 10 cm of extra height increases earnings by 13%.
- The IV point estimate is lower at 9% and not statistically significant.

Abstract

We use the Young Finns Study (N=~2000) on the measured height linked to register-based long-term labor market outcomes. The data contain six age cohorts (ages 3, 6, 9, 12, 15 and 18, in 1980) with the average age of 31.7, in 2001, and with the female share of 54.7. We find that taller people earn higher earnings according to the ordinary least squares (OLS) estimation. The OLS models show that 10 cm of extra height is associated with 13% higher earnings. We use Mendelian randomization, with the genetic score as an instrumental variable (IV) for height to account for potential confounders that are related to socioeconomic background, early life conditions and parental investments, which are otherwise very difficult to fully account for when using covariates in observational studies. The IV point estimate is much lower and not statistically significant, suggesting that the OLS estimation provides an upward biased estimate for the height premium. Our results show the potential value of using genetic information to gain new insights into the determinants of long-term labor market success.

Keywords: Height; Stature; Height premium; Earnings; Employment

Introduction

Taller people reap higher earnings. This empirical finding has been documented in several studies. There are three main explanations for the labor market premium in height (e.g., Sargent and Blanchflower, 1994; Judge et al., 2004; Persico et al., 2004; Case and Paxson, 2008; Tao, 2014; Sohn, 2015; Yamamura et al., 2015).¹ First, height is associated with cognitive skills (Case and Paxson, 2008). Second, non-cognitive skills, such as social skills, may play a role in the height premium (Persico et al., 2004). Based on these two explanations, height is related to other individual qualities, such as cognitive or non-cognitive skills. Third, there may also be other social explanations for the height premium, such as discrimination against short people in the labor market (e.g., Cinnirella and Winter, 2009) as a form of social-perceptual bias by which tall individuals are perceived to have more positive qualities irrespective of their true qualities (Hamstra, 2014).²

Most empirical studies treat height as an exogenous variable when examining the link between height and labor market outcomes.³ However, there may be important confounders, such as socioeconomic background, early life conditions and parental investments, that influence both a person's height and subsequent labor market outcomes. It is challenging to adequately account for the combined effect of these factors when

¹ Hübler (2016) summarizes the relevant literature.

² There are also empirical studies that point to the role of muscular strength as an explanation for the height premium (Lundborg et al., 2014). Böckerman et al. (2010) find only limited evidence for this view in the Finnish setting.

³ Case and Paxson (2006), and Vogl (2014) treat height as endogenously determined. Height is influenced by childhood nutrition, childhood environment and the prevalence of childhood diseases.

using the covariates that are available in observational studies. For example, parental investments are notoriously difficult to comprehensively and accurately measure.⁴ Thus, the causal effect of height on earnings and labor market success largely remains an open question.

The literature has pursued two approaches to address causal effects between height and earnings. First, some empirical studies have used a twin design. With twin data, it is possible to eliminate shared environmental factors, such as the family background, neighborhood and peer effects, and genetic factors (e.g., Böckerman and Vainiomäki, 2013). Second, two earlier studies used genetic instruments for height (von Hinke Kessler Scholder et al., 2013; Tyrrell et al., 2016).⁵ Genetic information could be helpful because genetic markers that are correlated with height should not directly affect the outcome variable of interest (i.e. earnings or employment). The specific instrument used in this paper is based on the findings in the genetics literature. There is substantial heritability for body height (Silventoinen et al., 2003). However, the contribution of individual genetic variants is modest. As a result, we used a genetic score with variants that genome-wide association studies (GWASs) have found to be significantly associated with height in extensive population samples (Allen et al., 2010), minimizing the weak instrument problem.

⁴ Under the assumption of time-invariant parental investments they can be accounted for using fixed effects in panel data if the explanatory variable is time-varying. However, because adult height is time-invariant, its effect would also be eliminated by using fixed effects. Nevertheless, twin data can be used as discussed in the next paragraph.

⁵ The usage of genetic instruments is known as “Mendelian randomization” in the medical literature. The basic idea of Mendelian randomization is that genetic factors are distributed randomly in the population so that genetic risk factors are independent of potential confounding factors. We explain this idea in detail in the next section of the paper.

We use administrative information on long-term labor market outcomes (earnings and labor market attachment). We chose this approach because cross-sectional measures of labor market outcomes are inaccurate proxies for individuals' lifetime labor market attachment and earnings (Böhlmark and Lindquist, 2006). Moreover, the use of the comprehensive register-based, long-term measures reduces measurement error from non-response and reporting biases.

Our contribution to the sparse empirical literature on the effects of height using genetic information builds on the fact that von Hinke Kessler Scholder et al. (2013) did not examine the labor market outcomes and that Tyrrell et al. (2016) used a self-reported categorical annual household income from a single year. In contrast, our paper uses linked data with administrative information and focuses on earnings that are a better measure of labor market success than annual household income that is confounded by social income transfers and spouse's income.

Methods

Mendelian randomization

Mendelian randomization refers to empirical studies that use genetic instruments to estimate the causal effects of exposure variables or traits in non-experimental (observational) data because it is often difficult or impossible to use randomized controlled trials (Tyrrell et al., 2016). The need for instrumentation arises from the presence of confounding factors that correlate with both the exposure and the outcome variable. This leads to bias in OLS estimation.

Figure 1 depicts the various effects and estimators in this setting. The IV or Wald estimator avoids the bias if the following conditions are fulfilled: (1) the genetic instrument (G) must correlate with the exposure variable (X), i.e. it must be informative; (2) the genetic instrument (G) must affect the outcome (Y) only through its effect on the exposure (X), i.e. the instrument must be exogenous; and (3) the instrument (G) and confounder (Z) must be independent, i.e. the genotype should not be associated with the confounding relations between (X) and (Y). The first condition is justified based on genome-wide association studies that provide evidence for the correlations between genotypes and exposure variables that in our research design is the height of persons, which we discuss in greater detail below. The second condition is essentially an assumption, which is strictly untestable, but some indirect evidence for its validity is provided below. The validity of the third condition is based on Mendel's second law (independent assortment), which states that genotypes are assigned randomly when passed across generations. This implies that in the population, or a representative sample of the population, genotypes are distributed independently of any confounding factors, which is what gives the method its name. It achieves the randomization of the exposure via "nature" rather than controlled trial.

Data sources and the sample

We use a longitudinal research design based on linked data. The data on height and genetic markers are from the Cardiovascular Young Finns Study (YFS).⁶ The YFS began in 1980, when 4,320 participants in six age cohorts (ages 3, 6, 9, 12, 15 and 18 years) were randomly chosen from five Finnish university regions using the national population

⁶ The YFS is described at <http://youngfinnsstudy.utu.fi/studydesign.html>

register (Raitakari et al., 2008). A total of 3,596 people participated in the study in 1980, and seven follow-up studies have been conducted; the most recent was in 2011/12.⁷

To obtain register information on labor market outcomes, we linked the YFS to the Finnish Longitudinal Employer-Employee Data (FLEED) of Statistics Finland (SF) using unique personal identifiers. The matching was exact, and there were no misreported ID codes. FLEED includes information on individuals' earnings and labor market attachment, which is taken directly from comprehensive administrative registers that are maintained by SF.

To account for key observable differences in the parental background, we linked the YFS to the Longitudinal Population Census (LPC) of SF from the year 1980. We used indicators for the parents' university-level education as family background variables.

Measures

We used both earnings and labor market attachment as labor market outcomes. Register-based, long-term earnings were measured as the average wage and salary earnings over the period 2001-2012. Labor market attachment was measured as the average employment years over the period 2001-2012.

⁷ Written informed consent was obtained from participants who were at least 9 years old and from the parents of younger participants. The research plan and data collection procedures were accepted by the participating universities review boards, and data collection was conducted according to WHO standards as well as the Helsinki Declaration.

Height was measured in 2001, when the participants were 24 to 39 years old. Height was collected in professional health examinations that were conducted at local health centers. Thus, the measure contains a minimal amount of measurement error.

We used the genetic score as an instrument for height. The score was calculated from data for the 180 single nucleotide polymorphisms (SNPs) that are significantly ($p < 5 \times 10^{-8}$) associated with height on the genome wide level according to Allen et al. (2010). A non-weighted genetic risk was calculated as the sum of genotyped risk alleles or imputed allele dosages carried by an individual.⁸ A Kernel density plot of the genetic score that is used in the IV estimations is presented in Appendix A2.

Genotyping was performed using a custom-made Illumina Human 670K BeadChip. Genotypes were called using the Illumina clustering algorithm (Teo et al., 2007). After quality control, there were 2442 samples and 546677 genotyped SNPs available for further analysis. Imputation was performed with MACH software and HapMap release 22 as a reference panel.

The genetic score has two major advantages. First, it is more powerful than any of the individual SNPs because it explains more variation in height. Second, it is more valid because it reduces the risk that any individual single nucleotide polymorphism will bias

⁸ More information on the SNPs and alleles used in calculating the risk score is presented in Appendix A1. We also used a weighted score in a robustness check below. A weighted genetic score was calculated as a sum of imputed allele dosages carried by an individual each multiplied by the effect size reported by Allen (2010). The weighted sum of effect alleles was divided by the mean effect size and transformed to the z-score. Effect alleles, effect sizes and the imputation quality of each variant used in the gene score calculation are reported in Appendix A1.

the IV estimates via an alternative biological pathway (pleiotropy) (Palmer et al., 2012). An alternative method to mitigate the bias caused by pleiotropy is to use individual SNPs with the Egger method (Bowden et al., 2015).⁹

A genetic risk score for blood pressure was calculated using 29 SNPs associated with systolic and/or diastolic blood pressure in a genome-wide association study (International Consortium for Blood Pressure Genome-Wide Association Studies, 2011). In another genome-wide association study, genetic risk scores for total cholesterol and triglycerides were calculated with 25 SNPs associated with the total cholesterol concentration and 24 SNPs associated with the triglyceride concentration (Teslovich et al., 2010). All genetic scores were calculated as the sum of genotyped risk alleles or imputed allele dosages carried by an individual.

A computerized cognitive testing battery (CANTAB[®]) was used to assess cognitive performance in 2026 participants in the latest YFS follow-up study in 2011. The cognitive test battery included the following five tests: 1) the motor screening test (MOT), which was used as a training/screening tool to indicate difficulties in test execution; 2) the paired associates learning test (PAL), which measures visual and episodic memory and visuospatial associative learning; 3) the spatial working memory test (SWM), which measures short-term and spatial working memory as well as problem solving; 4) the reaction time test (RTI), which measures reaction, movement speed and attention; and 5) the rapid visual information test (RVP), which measures visual processing, recognition and sustained attention. Each test produced several variables.

⁹ We use the standard econometric methods of economics because the relevant literature (in economics) has used similar methods also. Otherwise it would be difficult to compare our estimation results to the earlier ones.

Principal component analysis was conducted to identify components accounting for the majority of the variation of the dataset (see Appendix A3).¹⁰ Principal component analyses were also performed separately for all individual tests. The first components resulting from these analyses were considered to represent cognitive performance related to the particular cognitive domain. The component for the motor screening test was excluded from further analysis, as it did not discriminate the subjects, which indicated a ceiling effect. Other components were normalized using the rank order normalization procedure, resulting in four variables, each with a mean of 0 and standard deviation (SD) of 1. All available data for each cognitive test were used in the analyses; therefore, the number of participants varied between the models (data on PAL and RTI tests were excluded for N=177 participants due to technical problems with the test equipment, while N=51 refused to participate in all or some of the tests). Detailed descriptions and validation of the cognitive data have been reported previously (Rovio et al., 2016).

Personality characteristics were assessed in 1983 for the four oldest cohorts using the Hunter-Wolf A-B Rating Scale (Wolf et al., 1982).¹¹ The Hunter-Wolf Rating scale consists of four components (Aggression, Leadership, Responsibility and Eagerness-Energy) that were measured with 3-8 items. Responses to the items were given on a 7-point scale (1='totally disagree' and 7='totally agree') (Jokela and Keltikangas-Järvinen, 2009; Hintsala et al., 2014).

¹⁰ We used SAS and the PROC FACTOR command in the analysis.

¹¹ The Young Finns Study does not contain information on Big Five personality traits that started to gain popularity in the later part of the 1980s. The Hunter-Wolf rating scale was designed to measure behavior. Thus, instead of intrinsic tendencies, our personality characteristics measure an individual's typical behavioral patterns that arguably have a significant role as a determinant of labor market success.

Statistical methods

There are two main elements in our analyses. First, we run OLS models to replicate standard observational studies.¹² Second, we estimate instrumental variable models wherein height, in 2001, is instrumented using the genetic score, providing the causal effects of height on the long-term labor market outcomes. We control for birth month, birth year effects and gender. These factors potentially impact earnings and employment, but they are predetermined with respect to our outcome variables and independent of the height risk factor.

We also include the parental education level to control for any intergenerational correlation of socioeconomic status (cf. Öberg, 2014). This accounts for possible omitted variable bias from the socioeconomic status of parents, which could affect the genetic risk factors, such as through assortative mating within educational groups. This would violate the independence assumption of the IV design, as argued in von Hinke et al. (2016). The problem can be corrected by conditioning on parental education, particularly the mother's education level, which varies significantly between high and low risk factor individuals (Table 3).

Results

Descriptive evidence indicates that taller people have substantially higher earnings (Table 1). The baseline OLS estimates (Table 2, Panel A) show that height is statistically and economically significantly associated with average earnings over the period 2001-

¹² We use Stata 14.1 (College Station, Texas) in all our analyses. The instrumental variables results are obtained using the IVREG2 command.

2012. The quantitative size of the estimate is considerable. The coefficient for height implies that 10 cm of extra height is associated with 13% higher earnings. The estimate is roughly comparable to the earlier estimates that have been reported using Finnish data (Böckerman et al., 2010; Böckerman and Vainiomäki, 2013). The OLS estimate for labor market attachment is also statistically significant. The quantitative size of the effect corresponds to four months of additional employment during the twelve years covered by the data.

Next, we turn to the preferred IV estimates (Table 2, Panel B). The genetic score is a powerful instrument for height. In the first stage of IV, the F statistic on the instrument is 188 in the earnings regression, which exceeds the minimum standard of $F=10$ suggested in Staiger and Stock (1997) by a wide margin. Based on McClellan et al. (1994), we divide our estimating sample into those with an above-average value and those with a below-average value for genetic score and evaluate whether the two groups significantly differ in their observable characteristics that conceivably correlate with the second-stage outcome (Table 3). It is impossible to prove the null hypothesis that the instrument is uncorrelated with the second-stage error term. However, the lack of correlation between the instrument and observed variables, as shown in Table 3, is consistent with the exclusion restriction.¹³

¹³ We also tested the exclusion restriction by dividing the sample into three ranges based on the gender-specific risk score. Analogously to the Goldfeld and Quandt (1965) test for homoscedasticity, we omitted a quarter of the central observations and used the individuals in the remaining lowest 3/8 and highest 3/8 to evaluate whether these groups significantly differed in their observable characteristics. We found significant differences (at least at the 10% level) in the following characteristics: height in 2001, visual and episodic memory, leadership, annual income (father), high education (mother) and blood pressure. Note that we control for these additional characteristics in Table 4.

The IV point estimate (Table 2, Panel B) for the effect of height on earnings is 9% per 10 cm of extra height, and the estimate is no longer statistically significant. The effect of height on average employment years over the period 2001-2012 is also much lower and insignificant in the IV estimation. The earlier Finnish evidence on the amount of the height premium is very limited. Böckerman and Vainiomäki (2013) report a 3.3% earnings effect per 1 cm for men and a 1.4% effect for women, but the effect remains significant only for women when using twin-differences and IV estimation. Using genetic instrument, Tyrrell et al. (2016, p. 6) report that "... a genetically determined 1 SD (6.3 cm) greater height was associated with a 0.05 (0.03 to 0.07) SD increase in annual household income." Comparing this effect to our result is difficult because their annual household income is measured using five categories and because they do not report the SD for it.

An important finding of our research setting is that the IV estimates suggest that the OLS estimates are potentially upward biased. For example, specific early life conditions and parental investments that are not accounted for in the OLS models may lead to both greater height through differences in the children's nutrition and health and thereby to better long-term labor market outcomes later in life. This unaccounted confounder may explain the higher point estimates for height in the OLS estimation.

A caveat to the finding of a null effect for height in the IV setting is that it may be a symptom of low power due to an insufficient sample size. Freeman et al. (2013) show that the sample size needed in Mendelian randomization study is inversely proportional to two factors: the variation in the exposure or observable trait explained by the genetic instrument, and the square of the true effect size. Genome-wide association studies usually report low shares of explained variation for genotypes; therefore, the minimum

sample size required for detecting effects, with given levels of significance and power, becomes much larger than that in a randomized controlled trial. Alternatively, the power to detect an effect with a given sample size becomes low.

Brion et al. (2013) show that the power for the Mendelian randomization study with one instrument (one SNP or multiple SNP predictor) can be expressed as a function of the non-centrality parameter (NCP) of the test for the 2SLS regression coefficient to be zero. They further show that the NCP depends on a number of parameters: the causal effect of X on Y, the OLS estimate of it, the proportion of the exposure variable explained by the genetic predictor, and the variances of the X and Y variables. They provide an online tool to perform such power calculations, which we have utilized.¹⁴ Using the OLS and IV estimates obtained above, the observed variances of the X and Y variables, and the explanatory power of a genetic instrument in our data (0.048, i.e. ~5%), the required sample size for a test of no causal relationship with 0.05 type I error and 0.7 power is approximately N=13500. Therefore, our result of no causal effect should be interpreted with caution, as it may reflect our inability to reject the null hypothesis due to the relatively small sample size.¹⁵ On the other hand, the nature of our outcome measure, register-based long-term earnings, should substantially decrease the measurement error and thereby improve the precision of estimation compared to e.g., Tyrrell et al. (2016).

Table 4 contains additional estimation results after controlling for other genetic risk scores and observable differences in cognitive skills and personality characteristics. Allen

¹⁴ The web tool is available at <http://glimmer.rstudio.com/kn3in/mRnd/>

¹⁵ There is an earlier study that uses IV design with the same data that reports statistically significant effects of obesity on labor market outcomes (Böckerman et al., 2016). This suggests that the sample size is not an important issue *per se*.

et al. (2010) identified five loci for which the height-associated variant was also correlated with variants associated with other traits and diseases, particularly bone mineral density, rheumatoid arthritis, type 1 diabetes, psoriasis and obesity. Although the use of the genetic risk score significantly reduces the risk that an individual gene variant will bias the results (Palmer et al., 2012), we also added the genetic risk scores for blood pressure, total cholesterol, and triglycerides to control for potential pleiotropy. The results remain intact after this adjustment.

The most interesting finding from Table 4 is that the relationship between height and labor market outcomes is not statistically significant in the OLS estimation after adding the measures for cognitive skills to the set of covariates. Although the decline in the point estimate is partly attributed to the reduction in the sample size, this result suggests that physical height may be a marker of beneficial circumstances for developing higher cognitive skills (cf. Case and Paxson, 2008). However, the measures of cognitive skills are only available for a subsample of the original YFS data, implying that we have to consider these results with caution. Accounting for personality characteristics does not significantly impact on the estimates of how height influences earnings and employment.¹⁶

The YFS dataset is quite small for estimating separate results for subsamples, but we estimated models by gender to identify potential differences. The OLS estimates for both men and women are almost identical (not reported). The IV estimates are not statistically significant. In particular, the estimate for men is also very close to zero. There is earlier empirical evidence from other countries showing that the height premium is larger for

¹⁶ Hübler (2013) has shown using German panel data that the height premium disappears after controlling for personality traits because tall persons are relatively risk tolerant compared to their shorter peers.

men (Hübler, 2016). Using a self-reported categorical annual household income from a single year and a genetic instrument, Tyrrell et al. (2016) report that the effect of height on annual household income is approximately twice as strong in men as in women. The previous Finnish OLS estimates also give larger effects for men, but they are available from data that contain a different age cohort (Böckerman et al., 2010; Böckerman and Vainiomäki, 2013). The institutional setting of the labor market is also potentially important. In particular, wage compression and women's high labor force participation rate in Finland in the six age cohorts that we examine may contribute to the small difference in the height premium between women and men.

We additionally estimated separate models for the recession years (2009-2012) and pre-recession period (2001-2008) to account for the possible business cycle variation in the height effects, as reported in Böckerman and Vainiomäki (2013). Our OLS estimates (not reported) show that the point estimate for the height effect almost doubles in the recession compared to the pre-recession period, but the IV estimates remain insignificant in both periods.

Our baseline IV estimates used non-weighted genetic score based on 180 SNPs. As a final robustness check, we use alternative genetic scores. First, the IV estimates based on the weighted genetic score based on 180 SNPs reported in Appendix A4 are similar compared to the ones that used non-weighted genetic score in Table 2. Second, there is empirical literature that identifies more relevant SNPs for height (Wood et al., 2014) compared with the 10% explained by the 180 SNPs included in the genetic score used in the baseline estimates of this study. A more recent GWAS identified altogether 697 SNPs at a genome-wide significant level that explained 16% of phenotypic variance. The effect sizes of the weighted height scores for these are 1.97 cm/SD(GRS) for 180 SNPs and

2.80 cm/SD(GRS) for 697 SNPs in the entire YFS population. For this reason, we have also used information on 697 SNPs to construct the genetic score.¹⁷ Genotype imputation was performed using SHAPEIT v1 (Delaneau, 2011) and IMPUTE2 (Howie, 2009) software and the 1000G Phase I Integrated Release Version 3 as a reference panel (1000 Genomes Project Consortium, 2010). The results using non-weighted and weighted score based on 697 SNPs are reported in Appendix A5. The IV estimates using the alternative genetic score are similar to our baseline estimates in Table 2.

Discussion

Using the genetic score as an instrument for measured height, we find that the IV point estimate for the height premium is lower than the OLS estimate and is not significantly different from zero. Taken at face value, this suggests that the OLS estimates for the quantitative size of the height premium in the literature may be upward biased. The use of the genetic score as an instrument for height accounts for potential confounders that are related to socioeconomic background, early life conditions and parental investments that have not been systematically accounted for in previously reported empirical literature.

¹⁷ The use of the 180 SNPs is not an important limitation *per se* because the most important SNPs with largest effect sizes are usually identified first. This implies that the use of additional SNPs would most likely not improve the power of the first stage of the IV regressions much. Because our data is relatively small additional SNPs with smaller effect size can even produce results that have the wrong signs compared to GWAS. A large set of genetic instrumental variables would also increase the possibility that some of the instruments are invalid due to pleiotropy (Bowden et al., 2015).

As the genetic score (used as an instrument for height) was not associated with earnings or employment while true height was, the associations between height and labor market outcomes are not causal effects. Instead, our results suggest that physical height is a marker of beneficial circumstances for developing higher cognitive skills during childhood or adolescence due to family background. The disappearance of the height effect by inclusion of cognitive skills as an explanatory variable supports this interpretation.

Our results are broadly consistent with those reported in a twin design for monozygotic (identical) twins by Böckerman and Vainiomäki (2013). The authors found an insignificant height premium for men after controlling for genetic differences between twins, which supports that the cross-sectional OLS results are driven by unobserved differences such as cognitive skills. For women, they found that the height premium prevails in earned income, but not in capital income, for identical twins. This suggests that discrimination is a potential explanation.

Our study has two limitations. First, the number of observations in the baseline estimations is ~2,000. A larger data set would be needed to provide more power in order to obtain more tightly estimated effects. Second, the Young Finns Study that we use in the estimations is not nationally representative with respect to the total population in Finland, which consists also of the older age cohorts. The data are representative only for the selected six age cohorts in 1980. Further studies are needed to confirm that the patterns prevail in all age cohorts and other institutional settings.

Acknowledgements

The Young Finns Study has been financially supported by the Academy of Finland: grants 286284 (for T.L.), 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117787 (Gendi), and 41071 (Skidi); 293120 (Strategic Research Council funding for the project Work, Inequality and Public Policy); the Social Insurance Institution of Finland; Kuopio, Tampere and Turku University Hospital Medical Funds (for T.L. X51001); Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation of Cardiovascular Research and Finnish Cultural Foundation; Tampere Tuberculosis Foundation and Emil Aaltonen Foundation. The Palkansaaja Foundation supported the use of linked data. Jutta Viinikainen and Jaakko Pehkonen acknowledge financial support from the Yrjö Jahnsson Foundation (grants 6664 and 6646). Mirka Hintsanen acknowledges financial support from the Academy of Finland (project 258578). We thank Leo-Pekka Lyytikäinen for help with the genetic data. The authors would like to thank two anonymous referees and Tomi Kyyrä for valuable comments and suggestions that have greatly improved the paper.

References

- Allen, H.L., Estrada, K., Lettre, G., et al. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*, 467, 832–838.
- Bowden, J., Smith, G.D., Burgess, S. (2015). Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression. *International Journal of Epidemiology*, 44, 512–525.

Brion, M-J A., Shakhbazov, K., Visscher, P.M. (2013). Calculating statistical power in Mendelian randomization studies. *International Journal of Epidemiology*, 42, 1497–1501.

Böckerman, P., Johansson, E., Kiiskinen, U., Heliövaara, M. (2010). The relationship between physical work and the height premium: Finnish evidence. *Economics and Human Biology*, 8, 414–420.

Böckerman, P., Vainiomäki, J. (2013). Stature and life-time labor market outcomes: Accounting for unobserved differences. *Labour Economics*, 24, 86–96.

Böckerman, P., Cawley, J., Viinikainen, J., et al. (2016). The effect of weight on labor market outcomes: An application of genetic instrumental variables. National Bureau of Economic Research, Working Paper No. 22200.

Böhlmark, A., Lindquist, M.J. (2006). Life-cycle variations in the association between current and lifetime income: replication and extension for Sweden. *Journal of Labor Economics*, 24, 879–900.

Case, A., Paxson, C. (2008). Stature and status: Height, ability, and labor market outcomes. *Journal of Political Economy*, 116, 499–532.

Cinnirella, F., Winter, J., 2009. Size matters! Body height and labor market discrimination: A cross-European analysis. CESifo Working Paper No. 2733.

Delaneau, O., Marchini, J., Zagury, J.F. (2012). A linear complexity phasing method for thousands of genomes. *Nature Methods*, 9, 179–181.

Freeman, G., Cowling, B.J., Schooling, C.M. (2013). Power and sample size calculations for Mendelian randomization studies using one genetic instrument. *International Journal of Epidemiology*, 42, 1157–1163.

Goldfeld, S.M., Quandt, R.E. (1965). Some tests for homoscedasticity. *Journal of the American statistical Association*, 60, 539–547.

Hamstra, M.R.W. (2014). ‘Big’ men: Male leaders’ height positively relates to followers’ perception of charisma. *Personality and Individual Differences*, 56, 190–192.

Hintsala, T., Jokela, M., Pulkki-Råback, L., Keltikangas-Järvinen, L. (2014). Age- and cohort-related variance of Type-A behavior over 24 years: the Young Finns Study. *International Journal of Behavioral Medicine*, 21, 927–935.

Howie, B.N., Donnelly, P., Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genetics*, 5(6), e1000529.

Hübler, O. (2013). Are tall people less risk averse than others? *Schmollers Jahrbuch: Journal of Applied Social Science Studies*, 133, 23–42.

Hübler, O. (2016). Height and wages. In: Komlos, J., Kelly, I.R. (Eds.): *The Oxford Handbook of Economics and Human Biology*. Oxford University Press, Oxford, pp. 501–520.

International Consortium for Blood Pressure Genome-Wide Association Studies (2011). Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature*, 478, 103–109.

Jokela, M., Keltikangas-Järvinen, L. (2009). Adolescent leadership and adulthood fertility: Revisiting the “central theoretical problem of human sociobiology”. *Journal of Personality*, 77, 213–229.

Judge, T.A., Cable, D.M. (2004). The effect of physical height on workplace success and income: Preliminary test of a theoretical model. *Journal of Applied Psychology*, 89, 428–441.

Lundborg, P., Nystedt, P., Rooth, D.O. (2014). Height and earnings: The role of cognitive and noncognitive skills. *Journal of Human Resources*, 49, 141–166.

McClellan, M., McNeil, B.J., Newhouse, J.P. (1994). Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables. *Journal of the American Medical Association*, 272, 859–866.

Palmer, T.M., Lawlor, D.A., Harbord, R.M., et al. (2012). Using multiple genetic variants as instrumental variables for modifiable risk factors. *Statistical Methods in Medical Research*, 21, 223–242.

Persico, N., Postlewaite, A., Silverman, D. (2004). The effect of adolescent experience on labor market outcomes: The case of height. *Journal of Political Economy*, 116, 1019–1053.

Raitakari, O.T., Juonala, M., Rönnemaa, T., et al. (2008). Cohort profile: the cardiovascular risk in Young Finns Study. *International Journal of Epidemiology*, 37, 1220–1226.

Rovio, S.P., Pahkala, K., Nevalainen, J., et al. (2016). Cognitive performance in young adulthood and midlife: Relations with age, sex, and education—the cardiovascular risk in Young Finns Study. *Neuropsychology*, 30, 532–542.

Sargent, J.D., Blanchflower, D.G. (1994). Obesity and stature in adolescence and earnings in young adulthood – analysis of a British birth cohort. *Archives of Pediatrics and Adolescent Medicine*, 148, 681–687.

Silventoinen, K., Sammalisto, S., Perola, M., et al. (2003). Heritability of adult body height: A comparative study of twin cohorts in eight countries. *Twin Research*, 6, 399–408.

Sohn, K. (2015). The height premium in Indonesia. *Economics and Human Biology*, 16, 1–15.

Staiger, D., Stock, J.H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, 65, 447–486.

Tao, H.L. (2014). Height, weight, and entry earnings of female graduates in Taiwan. *Economics and Human Biology*, 13, 85–98.

Teo, Y.Y., Inouye, M., Small, K.S., et al. (2007). A genotype calling algorithm for the Illumina BeadArray platform. *Bioinformatics*, 23, 2741–2746.

Teslovich, T.M., Musunuru, K., Smith, A.V., et al. (2010). Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*, 466, 707–713.

Tyrrell, J., Jones, S.E., Beaumont, R., et al. (2016). Height, body mass index, and socioeconomic status: Mendelian randomisation study in UK Biobank. *British Medical Journal*, 352, i582.

Vogl, T.S. (2014). Height, skills, and labor market outcomes in Mexico. *Journal of Development Economics*, 107, 84–96.

von Hinke Kessler Scholder, S., Smith, G.D., Lawlor, D.A., et al. (2013). Child height, health and human capital: Evidence using genetic markers. *European Economic Review*, 57, 1–22.

von Hinke, S., Smith, G.D., Lawlor, D.A., et al. (2016). Genetic markers as instrumental variables. *Journal of Health Economics*, 45, 131–148.

Wolf, T.M., Sklov, M.C., Wenzl P.A., et al. (1982). Validation of a measure of Type A behavior pattern in children: Bogalusa Heart Study. *Child Development*, 53, 136–135.

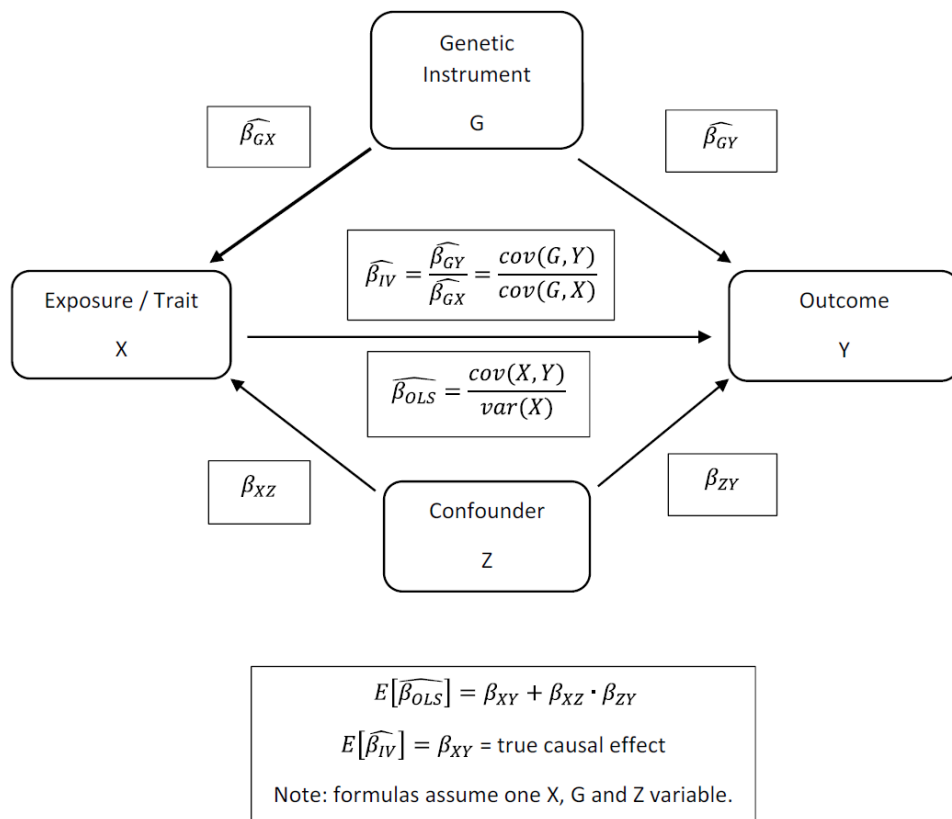
Wood, A.R., Esko, T., Yang, J., et al. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics*, 46, 1173–1186.

Yamamura, E., Smyth, R., Zhang, Y. (2015). Decomposing the effect of height on income in China: The role of market and political channels. *Economics and Human Biology*, 19, 62–74.

Öberg, S. (2014). Long-term changes of socioeconomic differences in height among young adult men in Southern Sweden, 1818–1968. *Economics and Human Biology*, 15, 140–152.

1000 Genomes Project Consortium. (2010). A map of human genome variation from population-scale sequencing. *Nature*, 467, 1061–1073.

Figure 1. Illustration of Mendelian randomization.



Appendix A2. Kernel density plot of the genetic score.

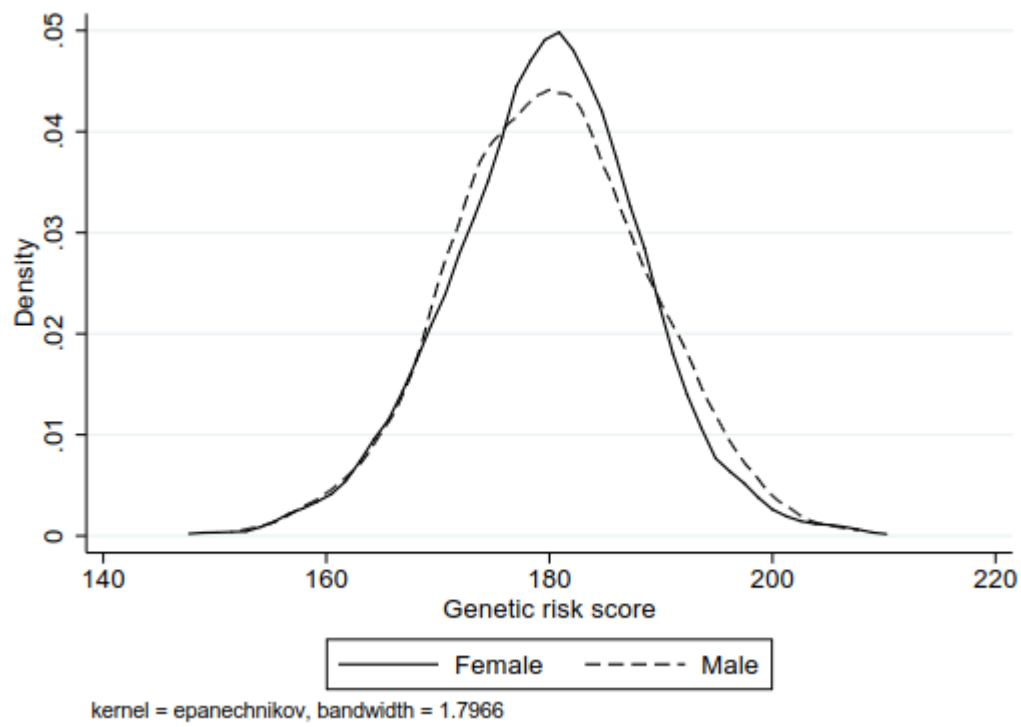


Table 1. Sample characteristics.

	Height below gender specific mean	Height above gender specific mean			
	Mean (SD)	Mean (SD)	Difference	t- statistics	N
Average earnings 2001-2012 (euros per year)	23391.17 (13185.86)	25663.80 (16405.44)	-2272.627	-3.399***	1982
Share of years employed, 2001-2012	0.858 (0.240)	0.878 (0.211)	-0.020	-1.979**	1982
Genetic score for height	177.836 (0.268)	181.816 (0.269)	-3.980	- 10.474***	1982
Married (2001)	0.438 (0.496)	0.458 (0.498)	-0.020	-0.903	1982
Cognitive skills (2011)					
Overall cognitive performance	-0.029 (1.012)	0.034 (0.973)	-0.063	-1.142	1302
Visual and episodic memory and visuospatial associative learning	0.001 (1.013)	0.019 (0.951)	-0.018	-0.335	1302
Reaction time	-0.019 (0.995)	0.064 (0.994)	-0.083	-1.511	1302
Rapid visual information processing	-0.013 (1.003)	0.087 (0.964)	-0.100	-1.829*	1302
Spatial working memory	-0.014 (0.970)	0.003 (0.980)	-0.017	-0.315	1302
Personality characteristics (1983)					
Aggression	3.850 (0.912)	3.806 (0.876)	0.044	0.803	1042
Leadership	4.221 (0.910)	4.322 (0.862)	-0.100	-1.830*	1042
Responsibility	4.861 (1.027)	4.846 (1.075)	0.015	0.227	1042
Eagerness-Energy	4.686 (0.705)	4.757 (0.706)	-0.071	-1.629	1042
Family background (1980)					
Annual income, mother (euros)	4497.34 (3519.50)	4771.26 (3481.42)	-273.91	-1.691*	1868
Annual income, father (euros)	8525.22 (5061.18)	9010.45 (6437.12)	-485.23	-1.811*	1868
High education, mother	0.068 (0.251)	0.078 (0.268)	-0.010	-0.865	1982
High education, father	0.094 (0.292)	0.111 (0.314)	-0.017	-1.259	1982
Other genetic risk scores					
Blood pressure	30.638	30.267	0.371	2.568**	1957

	(3.139)	(3.247)			
Total cholesterol	27.460 (3.123)	27.414 (3.057)	0.046	0.327	1957
Triglycerides	26.301 (2.860)	25.958 (2.900)	0.343	2.634***	1957

Table 2. Baseline results.

	Panel A: OLS		Panel B: IV	
	Coefficient (95% CI)	t-statistics	Coefficient (95% CI)	t-statistics
Log of average earnings, 2001-2012				
Height	0.013 (0.006, 0.020)***	3.91	0.009 (-0.012, 0.030)	0.84
F-statistics	..		187.71	
Mean outcome	9.867		9.867	
N	1982		1982	
Share of years employed, 2001-2012				
Height	0.003 (0.001, 0.004)***	3.12	0.001 (-0.004, 0.006)	0.33
F-statistics	..		187.71	
Mean outcome	0.868		0.868	
N	1982		1982	

Notes: Earnings are measured as the log of average earnings over the period of 2001-2012. Employment is measured as the average share of employment years over the period of 2001-2012. The mean values for the dependent variables are reported. Height was measured in 2001. All models include controls for the birth month, birth year effects, gender and parental education (1980). The instrument used in the IV models is the genetic score for height based on genetic markers. Angrist-Pischke multivariate F-tests of the excluded instrument are reported for the IV models. 95% confidence intervals based on heteroscedasticity-robust standard errors are reported in parentheses: significant at *10% **5% and ***1% levels.

Table 3. Comparison of observables by instrument value.

	Below gender specific average height score	Above gender specific average height score	Difference	t-statistics	N
Average earnings 2001-2012 (euros per year)	24389.03 (14468.18)	24661.27 (15354.84)	-272.239	-0.406	1982
Share of years employed 2001-2012	0.862 (0.239)	0.875 (0.213)	-0.013	-1.306	1982
Height in cm (2001)	170.684 (8.901)	173.661 (9.061)	-2.977	-7.375****	1982
Married (2001)	0.436 (0.496)	0.459 (0.499)	-0.023	-1.028	1982
Cognitive skills (2011)					
Overall cognitive performance	-0.029 (1.014)	0.034 (0.972)	-0.064	-1.154	1302
Visual and episodic memory and visuospatial associative learning	-0.034 (1.011)	0.053 (0.951)	-0.087	-1.606	1302
Reaction time	0.001 (0.999)	0.045 (0.992)	-0.043	-0.787	1302
Rapid visual information processing	0.049 (0.976)	0.027 (0.993)	0.022	0.395	1302
Spatial working memory	-0.017 (0.977)	0.007 (0.974)	-0.025	-0.453	1302
Personality characteristics (1983)					
Aggression	3.826 (0.900)	3.830 (0.889)	-0.004	-0.075	1042
Leadership	4.253 (0.914)	4.289 (0.862)	-0.036	-0.655	1042
Responsibility	4.826 (1.060)	4.879 (1.043)	-0.053	-0.809	1042
Eagerness-Energy	4.733 (0.712)	4.711 (0.701)	0.022	0.500	1042
Family background (1980)					
Annual income, mother (euros)	4599.33 (3601.71)	4667.80 (3404.99)	-68.467	-0.422	1868
Annual income, father (euros)	8548.13 (5873.54)	8979.62 (5709.11)	-431.495	-1.610	1868
High education, mother	0.061 (0.239)	0.084 (0.278)	-0.024	-2.043**	1982
High education, father	0.099 (0.298)	0.106 (0.308)	-0.008	-0.557	1982
Other genetic risk scores					
Blood pressure	30.537 (3.192)	30.371 (3.203)	0.165	1.144	1957
Total cholesterol	27.498 (3.137)	27.379 (3.043)	0.119	0.849	1957
Triglycerides	26.168 (2.879)	26.094 (2.892)	0.074	0.570	1957

Notes: Significant at *10% **5% and ***1% levels.

Table 4. Additional results, controlling for cognitive skills and personality characteristics.

	Panel A: OLS			Panel B: IV		
	Adding other genetic risk scores as a control	Adding cognitive skills as a control	Adding personality characteristics as a control	Adding other genetic risk scores	Adding cognitive skills as a control	Adding personality characteristics as a control
	Coef (95% CI)	Coef (95% CI)	Coef (95% CI)	Coef (95% CI)	Coef (95% CI)	Coef (95% CI)
Log of average earnings, 2001-2012						
Height	0.013 (0.007, 0.020)***	0.004 (-0.004, 0.011)	0.014 (0.005, 0.023)***	0.009 (-0.012, 0.030)	0.007 (-0.017, 0.030)	0.007 (-0.019, 0.033)
F-statistics	180.54	135.69	114.97
N	1957	1302	1042	1957	1302	1042
Share of years employed, 2001-2012						
Height	0.003 (0.001, 0.005)***	0.000 (-0.002, 0.002)	0.003 (0.001, 0.006)***	0.001 (-0.004, 0.007)	0.000 (-0.006, 0.006)	0.001 (-0.006, 0.007)
F-statistics	180.54	135.69	114.97
N	1957	1302	1042	1957	1302	1042

Notes: The effects of additional controls are not reported. Earnings are measured as the log of average earnings over the period of 2001-2012. Employment is measured as the average share of employment years over the period of 2001-2012. Height was measured in 2001. All models include controls for the birth month, birth year effects, gender and parental education (1980). The instrument used in the IV models is the genetic score for height based on genetic markers. Angrist-Pischke multivariate F-tests of the excluded instrument are reported for the IV models. 95% confidence intervals based on heteroscedasticity-robust standard errors are reported in parentheses: Significant at *10% **5% and ***1% levels.

Appendix A1. SNPs and the genotyped risk alleles used for the genetic score of height.

SNP id	Effect allele	Beta	Imputation quality (info)
rs425277	T	0.022	0.996
rs2284746	C	-0.04	0.9859
rs1738475	C	0.025	0.9987
rs4601530	T	-0.028	1
rs7532866	A	0.021	0.9793
rs2154319	T	-0.03	0.9161
rs17391694	T	0.042	0.995
rs6699417	T	0.021	1
rs10874746	T	-0.024	0.9969
rs9428104	A	-0.041	0.9886
rs11205277	A	-0.046	0.9939
rs17346452	T	-0.04	0.9861
rs1325598	A	-0.022	0.9968
rs1046934	A	-0.044	1
rs10863936	A	-0.021	1
rs6684205	A	-0.028	0.9937
rs11118346	T	-0.025	0.9893
rs10799445	A	0.032	0.9863
rs4665736	T	0.029	0.9337
rs6714546	A	-0.026	0.9911
rs17511102	A	-0.06	0.9565
rs2341459	T	0.025	0.9998
rs12474201	A	0.028	0.9939
rs3791675	T	-0.053	0.9959
rs11684404	T	-0.028	0.9987
rs7567288	T	-0.032	0.9832
rs7567851	C	0.037	1
rs1351164	T	0.034	0.995
rs12470505	T	0.041	0.9975
rs2629046	T	0.024	0.9991
rs2580816	T	-0.045	0.9983
rs12694997	A	-0.024	0.9997
rs2597513	T	-0.036	0.9998
rs13088462	T	-0.052	0.9976
rs2336725	T	-0.027	0.9906
rs9835332	C	-0.026	1
rs17806888	T	0.036	0.9641
rs9863706	T	-0.031	0.9944
rs6439167	T	-0.034	0.9978
rs9844666	A	-0.024	0.9985
rs724016	A	-0.07	0.9984
rs572169	T	0.033	0.9989

rs720390	A	0.029	0.9953
rs2247341	A	0.025	0.991
rs6449353	T	0.075	0.9553
rs17081935	T	0.03	0.9986
rs7697556	T	0.028	0.9835
rs788867	T	-0.043	0.987
rs10010325	A	0.024	0.9998
rs7689420	T	-0.073	0.9995
rs955748	A	-0.023	1
rs1173727	T	0.034	0.999
rs11958779	A	-0.027	0.9997
rs10037512	T	0.032	0.9976
rs13177718	T	-0.04	0.9998
rs1582931	A	-0.023	0.8666
rs274546	A	-0.029	0.9986
rs526896	T	0.03	0.9177
rs4282339	A	-0.036	0.9956
rs12153391	A	-0.03	0.9429
rs889014	T	-0.03	0.9858
rs422421	T	-0.031	0.9362
rs6879260	T	-0.022	0.9283
rs3812163	A	-0.036	0.9722
rs1047014	T	-0.032	0.996
rs806794	A	0.052	0.9911
rs3129109	T	-0.032	0.9701
rs2256183	A	0.04	0.9995
rs6457620	C	-0.029	0.9991
rs2780226	T	-0.076	0.9876
rs6457821	A	-0.104	0.9181
rs9472414	A	-0.026	1
rs9360921	T	-0.042	0.9999
rs310405	A	0.026	0.9897
rs7759938	T	-0.045	0.9878
rs1046943	A	0.02	0.9972
rs961764	C	-0.024	0.9725
rs1490384	T	0.034	0.9748
rs6569648	T	-0.04	0.9992
rs7763064	A	-0.048	0.9938
rs543650	T	-0.034	0.9978
rs9456307	A	-0.048	0.8046
rs798489	T	-0.048	0.9921
rs4470914	T	0.029	0.9998
rs12534093	A	-0.034	0.9638
rs1708299	A	0.04	0.9873
rs6959212	T	-0.024	0.9834
rs42235	T	0.057	0.966
rs822552	C	-0.025	0.8303

rs2110001	C	-0.031	0.9028
rs1013209	T	-0.025	0.9704
rs7460090	T	0.058	0.993
rs6473015	A	-0.029	0.9998
rs6470764	T	-0.05	0.9929
rs12680655	C	0.028	0.9998
rs7864648	T	0.022	0.9983
rs11144688	A	-0.049	0.4495
rs7853377	A	-0.024	0.9623
rs8181166	C	0.026	0.9772
rs2778031	T	0.031	0.9584
rs9969804	A	0.03	1
rs1257763	A	0.069	0.5821
rs473902	T	0.065	0.5173
rs7027110	A	0.031	1
rs1468758	T	-0.026	0.9938
rs751543	T	0.026	0.8534
rs7466269	A	0.032	0.9997
rs7849585	T	0.029	0.9933
rs7909670	T	-0.021	0.9985
rs2145998	A	-0.026	0.9997
rs11599750	T	-0.028	0.9997
rs2237886	T	0.046	0.9947
rs7926971	A	-0.023	0.9948
rs1330	T	0.022	0.9995
rs10838801	A	-0.027	0.9935
rs1814175	T	0.022	0.9703
rs5017948	A	0.027	0.8788
rs3782089	T	-0.058	0.9562
rs7112925	T	-0.023	0.9955
rs634552	T	0.039	0.9934
rs494459	T	0.02	1
rs654723	A	0.025	0.8502
rs2856321	A	-0.029	0.9993
rs10770705	A	0.033	0.9998
rs2638953	C	0.032	1
rs2066807	C	-0.054	0.8978
rs1351394	T	0.06	0.9851
rs10748128	T	0.038	0.9985
rs11107116	T	0.052	1
rs7971536	A	-0.028	0.924
rs11830103	A	-0.035	0.9901
rs7332115	T	-0.023	1
rs3118905	A	-0.056	0.9996
rs7319045	A	0.025	0.9681
rs1950500	T	0.034	0.9999
rs2093210	T	-0.032	0.9512

rs1570106	T	-0.026	0.9985
rs862034	A	-0.028	0.9888
rs7155279	T	-0.024	0.9621
rs16964211	A	-0.05	0.9998
rs7178424	T	-0.021	0.9999
rs10152591	A	0.041	0.9997
rs12902421	T	-0.062	0.9962
rs5742915	T	-0.031	0.9325
rs11259936	A	-0.044	1
rs16942341	T	-0.13	0.968
rs2871865	C	0.057	0.9402
rs4965598	T	-0.028	0.989
rs11648796	A	-0.034	0.9535
rs26868	A	0.034	0.9899
rs1659127	A	0.027	0.9533
rs8052560	A	0.029	0.7692
rs4640244	A	0.024	0.9952
rs3110496	A	-0.022	0.9939
rs3764419	A	-0.035	0.9994
rs17780086	A	0.028	0.9857
rs1043515	A	-0.023	0.9968
rs4986172	T	-0.032	0.9763
rs2072153	C	0.021	1
rs4605213	C	0.021	0.9571
rs227724	A	-0.03	0.9471
rs2079795	T	0.04	0.9995
rs2665838	C	-0.042	0.9712
rs11867479	T	0.025	0.9982
rs4800452	T	0.051	0.9995
rs9967417	C	-0.038	0.9695
rs17782313	T	-0.028	0.9995
rs12982744	C	-0.03	0.9753
rs7507204	C	0.036	0.9457
rs891088	A	-0.029	1
rs4072910	C	-0.031	0.7501
rs2279008	T	0.025	0.9958
rs17318596	A	0.032	0.9993
rs1741344	T	-0.023	0.997
rs2145272	A	-0.039	0.9976
rs7274811	T	-0.041	0.9961
rs143384	A	-0.063	0.8833
rs237743	A	0.041	1
rs2834442	A	0.026	0.9829
rs4821083	T	0.031	0.9997

Note: All SNPs are imputed with MACH software and HapMap release 22 as a reference panel.

Appendix A3. Eigenvalues and proportions of variance explained by principal components for cognitive performance used in the study.

	Eigenvalue	Explained variance
Component for Paired Associates Learning test (PAL)	7.72	0.55
Component for Spatial Working Memory test (SWM)	5.06	0.42
Component for Reaction Time test (RTI)	2.29	0.38
Component for Rapid Visual Information test (RVP)	5.19	0.74

Notes: Principal component analyses were performed separately for all individual tests. The first components resulting from these analyses were considered to represent cognitive performance related to the particular cognitive domain.

Appendix A4. IV results; weighted genetic score based on 180 SNPs.

	Coefficient (95% CI)	t-statistics
Log of average earnings, 2001-2012		
Height	0.006 (-0.013, 0.024)	0.59
F-statistics	212.82	
Mean outcome	9.867	
N	1981	
Share of years employed, 2001-2012		
Height	0.000 (-0.005, 0.005)	-0.02
F-statistics	212.82	
Mean outcome	0.868	
N	1981	

Notes: The effects of additional controls are not reported. Earnings are measured as the log of average earnings over the period of 2001-2012. Employment is measured as the average share of employment years over the period of 2001-2012. Height was measured in 2001. All models include controls for the birth month, birth year effects, gender and parental education (1980). The instrument used in the IV models is the genetic score for height based on genetic markers. Angrist-Pischke multivariate F-tests of the excluded instrument are reported for the IV models. 95% confidence intervals based on heteroscedasticity-robust standard errors are reported in parentheses: Significant at *10% **5% and ***1% levels.

Appendix A5. IV results; genetic score based on 697 SNPs.

	Panel A: IV, non-weighted score		Panel B: IV, weighted score	
	Coefficient (95% CI)	t-statistics	Coefficient (95% CI)	t-statistics
Log of average earnings, 2001-2012				
Height	0.006 (-0.007, 0.019)	0.94	0.003 (-0.009, 0.015)	0.50
F-statistics	453.57		488.91	
Mean outcome	9.867		9.867	
N	1982		1982	
Share of years employed, 2001-2012				
Height	0.001 (-0.003, 0.004)	0.52	-0.000 (-0.004, 0.003)	-0.20
F-statistics	453.57		488.91	
Mean outcome	0.868		0.868	
N	1982		1982	

Notes: The effects of additional controls are not reported. Earnings are measured as the log of average earnings over the period of 2001-2012. Employment is measured as the average share of employment years over the period of 2001-2012. Height was measured in 2001. All models include controls for the birth month, birth year effects, gender and parental education (1980). The instrument used in the IV models is the genetic score for height based on genetic markers. Angrist-Pischke multivariate F-tests of the excluded instrument are reported for the IV models. 95% confidence intervals based on heteroscedasticity-robust standard errors are reported in parentheses: Significant at *10% **5% and ***1% levels.