The Effect of BMI and Height on Socio-Economic Status
Evidence from the Health and Retirement Study

Andrea Pogliano

Advisor: D. Muslimova

June 2019

Abstract

Socio-economic status (SES) is correlated with better health and longer lifetime. The determinants of socio-economic status are, however, unclear, as there are many observed and unobserved factors feeding back and forth and causing reverse causality. This paper aims to verify whether physical appearance, in the form of height and body mass index (BMI), has a causal effect on SES. Possible explanations of the phenomenon include discrimination on the labor market as well as health problems. To investigate this relationship, I use insights from genetic studies and the Mendelian randomization approach. In this, genes work as instrumental variables to mitigate the issues of reverse causality and omitted variables bias. Analyzing data from the Health and Retirement Study (HRS) focused on the elderly American population, I find significant causal effects of BMI for four out of six proxies of SES while height influences income only. More specifically, I find that individuals with lower stature and higher BMI are at disadvantage for what concerns education and labor market outcomes.

Contents

Introduction

Insight from health economics reveals that heterogeneity in socio-economic status (i.e. the social standing of an individual measured as a combination of education, income and occupation) might lead to differences in health outcomes and in health service consumption (Smith, 2007; Vyas & Kumaranayake, 2006). Scholars explain this relationship through different hypotheses (the direct income hypothesis, the productive time hypothesis, etc.) but all seem to agree that socio-economic status (SES) has a positive effect on the health of an individual (Adler et al., 1994). Higher SES, moreover, is associated with higher life expectancy (Lin, Rogot, Johnson, Sorlie, & Arias, 2003). Marmot (2017), for example, found a difference of 20 years in male life expectancy between the richest and the poorest neighborhoods of Baltimore (US).

Based on this evidence, SES seems an important factor that affects the health and the life expectancy of citizens, this notion is known in literature as the socio-economic gradient of health. Due to the impact of SES on health, many policies in the healthcare sector take it into account. An example is provided by risk equalization: in the Netherlands, one of the risk factors included in risk equalization payments is indeed SES. Given that SES is such a vital element for current healthcare policies, a thorough understanding of its nature and of its determinants is beneficial. Unraveling the factors that affect SES will allow to analyze possible causes of disparity in health and healthcare consumption, enabling policy-makers to address the problem at its source.

In this regard, past research has been focusing on the impact that body appearance and physical attributes might have on SES (Roszell, Kennedy, & Grabb, 1989). Recent literature identified two physical variables that are associated with SES, namely body mass index (BMI) and height. For both variables, most research aims to explain the correlation through the causal effect of SES on an individual's weight and height (Jansen & Hazebroek-

Kampschreur, 1997; Murasko, 2009). Lower income may impact obesity through different mechanisms: lack of access to healthy food, unhealthy lifestyle, as well as psychosocial factors that derive from relative deprivation (Kim & von dem Knesebeck, 2018). Moreover, low SES in childhood might have a negative impact on growth, this leading to lower height (Tyrrell et al., 2016).

Alongside these theories, scholars have also started investigating the possibility of reverse causality: BMI and height could be the causes of a certain SES rather than its effect (Sobal & Stunkard, 1989). The explanation for this could be found in social stigma: obese and shorter people may appear as less productive and be discriminated on the labor market (Bann, Johnson, Li, Kuh, & Hardy, 2017; Kim & von dem Knesebeck, 2018). Another possible explanation could be that high BMI leads to lower health and this might in turn have a negative effect on the time dedicated to work and education. This paper aims to further understand the direction of this relationship, the following research question is thus presented:

"To what extent do body mass index (BMI) and height have a causal effect on socio-economic status (SES) indicators in a sample of American citizens?"

Understanding the causality in this relationship is of high social and scientific relevance. First, it will be beneficial in shaping policies aimed at promoting greater health among citizens. In case of a causal effect of BMI and height on SES, policies could either aim at counteracting social stigma and labor market discrimination or to fight obesity by taxing unhealthy food. In this sense, incentives to contrast poor parental food and lifestyle choices could be effective in reducing obesity and fostering a healthy growth. Determining whether the causal effect is present is therefore socially relevant.

In terms of scientific relevance, this paper aims at expanding the existing literature on the effect of BMI and height on SES. To do so, I use Mendelian

randomization, a tool that has just recently gained relevance in economic studies. The analysis follows the framework of Tyrrell et al. (2016), who analyzed the effect of BMI and height on SES using genetic data coming from the UK Biobank. The key contributions of this research is twofold. First, I employ data coming from the US. So far, the main studies on this topic were conducted with data coming from English databases (Bann et al., 2017; Tyrrell et al., 2016), analyzing the relationship with data from another country allows to check for the external validity of past results. Second, I use different proxies for SES that were not analyzed before, these include individual total wealth and position with respect to a poverty threshold.

After the introduction, I include the theoretical framework, in which I critically assess the main completed research on the effect of BMI and height on SES. I then present the methodology and discuss the technique of Mendelian randomization and its assumptions. Afterwards, I describe the data sampling process and include descriptive statics as well as an explanation of the transformations applied to the data. I then present the results and interpret them. Finally, I discuss the main findings and address the limitations of this paper.

Theoretical Framework

BMI and SES

For what concerns BMI, scholars have extensively researched the economic consequences of obesity, assessing whether obesity lowers wages or reduces employment and education (Atella, Pace, & Vuri, 2008; Cawley, 2007; Conley & Glauber, 2006; Morris, 2007). Obesity could result in worse SES for several reasons. First, obesity is associated with many diseases, such as hypertension, coronary heart disease, and diabetes (Greve, 2008). The worsened health deriving from such diseases might lead to lower productivity and thus to lower wages (Hu, 2008). Second, obesity might lead to discrimination on the labor market, as studies suggest that obese individuals are more

likely to be perceived as lazy and unable to self-control (Puhl & Latner, 2007; Rooth, 2009). Rooth (2007) found that discrimination against obese people in the hiring process is more common in occupations with more customer contact, this suggesting that obese people may appear unfit to convey a proper impression of the business. Third, these stereotypes are often internalized by the individuals themselves, this leading to self-stigma and psychological problems, which might in turn lower productivity (Kim & von dem Knesebeck, 2018). Finally, the impact of obesity on education follows a similar mechanism, with obese students possibly being self-aware of their issue and limited in their productivity and soft skills.

Beyond the possible mechanisms, different estimations of the effect are also present. The magnitude of the effect of BMI on SES varies widely from research to research, due to the different econometric techniques used to account for the issues of omitted variable bias and reverse causality. Many researchers used the weight of a relative as an instrument, exploiting the fact that there is a strong heritable component of BMI. Brunello and d'Hombres (2007) found that a 10% increase in BMI leads to a decrease in wages of 3.27% and 1.86% for males and females respectively. On the other side, Cawley (2007) found an effect on women only, with a difference of two standard deviation from the mean BMI leading to a decrease in wage of 9%. Using the weight of a relative as an instrument, however, rises concerns about possible confounders, as there could be variables that affect both the BMI of the relative and the outcome variable. For this reason, Mendelian randomization has been used in recent research. As later explained, Mendelian randomization is a technique which uses genes variants as instrumental variables and exploits their exogeneity to account for confounders. Among the main studies, Tyrrell et al. (2016) found that a standard deviation from the mean in BMI leads to a decrease of £210 and £1890 in household annual income for men and women respectively. This paper aims to confirm these results and extend them to a different dataset. Given past research, it seems reasonable to expect a negative effect of BMI on different measures of SES, I thus formulate the following hypothesis:

H1. There is a negative causal relationship between an individual's BMI and SES

Height and SES

Research on the effect of height on SES uses similar reasoning as the one for BMI. Also in this case, a possible mechanism through which height affects SES relies on social consideration. Scholars have underlined how factors such as social dominance (Hensley, 1993) and self-esteem (Young & French, 1996) might help taller people in the labor market. To this extent, Judge and Cable (2004) found that height is correlated to measures of social esteem, leader emergence and performance. Frieze, Olson, and Good (1990) argued that, due to these stereotypes, short people might be considered less favorably and thus have a disadvantage when arranging interpersonal dealings. In this regard, Rosenberg (2009) claimed that heightism, i.e. the prejudice or discrimination of a person based on his height, is a common phenomenon in the labor market. Evidence for this was also found by Cinnirella and Winter (2009), who determined a link between height and salaries for employed workers and not for the self-employed, this possibly implying employer discrimination. Finally, another possible mechanism by which height influences SES is introduced by Persico, Postlewaite, and Silverman (2004), who argued that the height premium is partially mediated through participation in high school sports and clubs, which helps to build productive human capital.

Also for height, the estimation of the effect varies widely across research and depend on the methods employed to account for reverse causality. Using sibling data from Sweden, Magnusson, Rasmussen, and Gyllensten (2006) found that men taller than 1.94 meters are twice more likely to pursue higher education as compared to men shorter than 1.65 m. Hübler (2009) found that on the labor market in Germany a 10-cm height increase was associated with a 2.5% wage increase for men, with a lower estimation for women. Using Mendelian randomization, Tyrrell et al. (2016) found that a standard devia-

tion increase from the mean height leads to 0.06 years more years of education completed as well as increasing the odds of working in a skilled profession. The literature suggests therefore a positive impact of height on education and labor market outcomes, to test this claim I formulate the following hypothesis:

H2. There is a positive causal relationship between an individual's height and SES

Methodology

Mendelian randomization

As mentioned, this paper relies on the methodology of Tyrrell et al. (2016) and uses the method of Mendelian randomization. Mendelian randomization is an approach that employs genetic variants as instrumental variables to analyze the causal effects of one trait (the 'exposure') on another (the 'outcome') (Brumpton et al., 2019). In this case, genetic scores associated with BMI and height are used as instrumental variables to investigate the effect of BMI and height on SES. Mendelian randomization has recently gained popularity due to decrease in the cost of genotyping and the consequent increase in the availability of genetic data. Mendelian randomization, moreover, can help alleviating the issues of omitted variables and reverse causality common to other methods (Brumpton et al., 2019). For Mendelian randomization to give an unbiased estimation, the three conditions of instrument variables need to be met.

First, the instrument needs to be relevant, or, in other words, the genetic score needs to be associated with the exposure; this assumption is met due to the predictive power of the genetic variants for BMI and height. A study conducted on the dataset in analysis found that the genetic scores for height and BMI can explain the 24.6% of the variance of height and 6.0% of the variance of BMI respectively (Yengo et al., 2018).

Second, the instrument must be independent, so that there are no confounders that affect both the instrument and the outcome. In other words, the height and BMI genetic scores should not be associated with confounding factors that could bias the associations between height/BMI and socioeconomic status. This assumption is reasonably met, as genes are per se exogenous. Recent research, however, has suggested the possibility of bias coming from dynasty effects and assortative mating (Brumpton et al., 2019; Hartwig, Davies, & Davey Smith, 2018). Dynasty effects imply that the genes of the parents are linked both to the genes of the offspring and the SES of the children during childhood, which in turn affects the SES of the individual in a later stage of life. Assortative mating occurs when individuals choose a partner for certain characteristic or, in other words, for certain genes they possess. This can lead to spurious genetic associations in the offspring and then result in biased estimations. To solve these issues, two approaches could be undertaken. Brumpton et al. (2019) suggest that a within-family design could be implemented as this would permit to control for family fixed effects and account for those variables that are caused by parents' genes and affect the living situation of the children. Alternatively, the genetic scores of the parents could be used as control variables (Hartwig et al., 2018). Unfortunately, the dataset used contains very limited information about the family of origins and therefore it is not possible to control for family fixed effects nor to use genetic data of the parents. The possible violation of this assumption remains therefore a major limitation of the paper and will be further discussed in the conclusion.

The third and final assumption, i.e. the exclusion restriction, specifies that the genes should affect the outcome entirely through the exposure. In this case, BMI/height genes should affect SES only through their effect on BMI and height. A concern about this assumption is also present, due to the possibility of horizontal pleiotropy. Horizontal pleiotropy occurs when the genetic variant affects traits outside of the pathway of interest that in turn have an impact on the outcome (Verbanck, Chen, Neale, & Do, 2018). In this

case, it might be that genes associated with BMI could also influence other characteristics (education propensity, etc.) that have an impact on SES. To investigate whether this is a problem in this research, I could undertake two approaches, following the argumentation of Böckerman et al. (2019). First, a qualitative study of the literature indicates that genes associated with BMI and height seem not to affect other traits that could in turn affect SES outcomes. More specifically, it was found that the genes have no effect on an individual's intelligence (Speliotes et al., 2010). Another way to investigate possible pleiotropy follows the idea of McClellan, McNeil, and Newhouse (1994) and makes use of a placebo test, whose procedure I explain later. In conclusion, given that the assumptions are met and considering the limitations, the method of Mendelian Randomization seems opportune to find a causal effect due to the elimination of most confounders, as represented in the causal pathway in Figure 1.
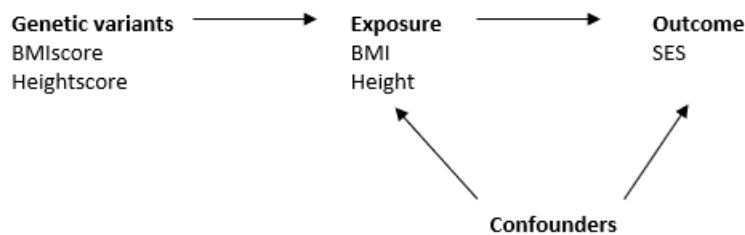
Figure 1: Causal Pathway

Implementation

To test the first hypothesis, I use six instrumental regression analyses. The exposure (independent) variable is BMI, the instrument is the genetic score related to BMI and the independent variables are six different proxies for SES. The proxies include years of education, whether the individual attended college (binary variable), individual earnings, total wealth, position with respect to a poverty threshold (a binary variable created by the Health and Retirement Study that combines income and household composition) and

labor force status (binary variable). Finally, I add control variables both in the first and second stage. The control variables are the age of the individual in months and ten principal components, which can be employed to adjust for population structure and potential outliers. The variable PC is therefore an array of 10 variables representing the principal components. Moreover, given the results of the Placebo test discussed later, I also include parental education as a control in the form of two variables, maternal and paternal education. The first stage of the regression model is the following:

$$\text{BMI}_i = \alpha_1 + \gamma_{\text{BMI}}\text{BMIscore}_i + \mu_1\text{Age}_i + \rho_1\textbf{PC}_i + \zeta_1\textbf{ParentalEducation}_i + \epsilon_i \tag{1}$$

Where BMIscore is the genetic score of an individual, γ_{BMI} its coefficient, α_1 the constant and ε_i the error term. Such an equation is used to find the predicted values for BMI used in the second stage:

$$Y_i = \delta_1 + \beta_{\text{BMI}}\widehat{\text{BMI}}_i + \mu_2\text{Age}_i + \rho_2\textbf{PC}_i + \zeta_2\textbf{ParentalEducation}_i + e_i \tag{2}$$

Where Y is the value of the different proxies, $\widehat{BMI}_i$ the predicted value of BMI, δ_1 the constant and e_i the error term. Note that in the case that Y is a binary variable, I use a linear probability model (LPM), as done by Norton and Han (2008). This makes the interpretation of the coefficients more immediate and does not meaningfully change the conclusions when compared to results from a Probit model (Norton & Han, 2008). Still, the coefficients should be interpreted with cautions as the LPM model might lead to predictions outside the 0-1 range. For the hypothesis not to be rejected, a negative sign for the coefficient β_{BMI} is expected.

To test the second hypothesis, the same approach is used. Like before, I employ six instrumental regression analyses with the same proxies for SES. The first stage of the regression model is:

$$\text{Height}_i = \alpha_2 + \gamma_{\text{Height}}\text{Heightscore}_i + \mu_3\text{Age}_i + \rho_3\mathbf{PC}_i + \zeta_3\mathbf{ParentalEducation}_i + f_i \tag{3}$$

Where Heightscore is the genetic score of an individual, γ_{height} its coefficient, α_2 the constant and f_i the error term. Such an equation is used to find the predicted values for height used in the second stage:

$$Y_i = \delta_2 + \beta_{\text{Height}}\widehat{\text{Height}}_i + \mu_4\text{Age}_i + \rho_4\mathbf{PC}_i + \zeta_4\mathbf{ParentalEducation}_i + t_i \tag{4}$$

As before, Y is the value of different proxies and a positive coefficient for β_{height} is expected, δ_2 will be the constant and t_i the error term. As before, when Y is a binary variable, I employ an LPM. Since past research found different estimation for different genders, I conduct also each regression separately for men and women.

As mentioned, I also include a placebo test to investigate the presence of pleiotropy and dynasty effects. I divide the sample into two groups, one with above-average genetic score and the other with below-average values. Afterwards, I test whether the two groups significantly differ in their observable characteristics (parents' education, cognition scores) that are likely to be correlated with SES or with parents' genes. Similarity for the two groups in the observed variables, although not proving it, would be consistent with the exclusion restriction and the independence assumption. In addition, I perform a regression with the genetic scores (dependent variables) and the aforementioned variables (independent variables). Insignificant coefficients would also provide support to the assumptions.

Data

Sampling

The paper uses data from the Health and Retirement Study (HRS), a longitudinal household survey conducted in the US, and combines two publicly available datasets, the RAND HRS Longitudinal File and the Polygenic Score Data. The RAND HRS Longitudinal File contains data on 12 survey sessions, or waves, completed by 37495 individuals between 1992 and 2014. In this paper, I use data from wave 10 (interviews recorded in 2010), as it contains the highest number of observations for the variables of interest. Moreover, I consider only individuals still active in the labor market at the time of the interview as, only for them, BMI and height could still influence income. Retired and disabled individuals are therefore removed from the sample, following the example of Tyrrell et al. (2016). Moreover, given that the genetic data employed is scaled on European descent, I use only observations from individuals of white/Caucasian race. Eventually, the sample includes 3156 observations. The variables of interest taken from the sample include information regarding BMI and height (exposure variables) as well as individual earnings, total wealth, labor force status, poverty threshold, total education and highest level of education reached (outcome variables). Note that the BMI and height at the time of education are not available in the dataset, this problem is solved by considering BMI and height in wave 10 as proxies for BMI and height throughout life, as done by Tyrrell et al. (2016). From the same dataset, other demographic information is collected, such as year of birth and gender, plus other variables needed for the placebo test (parents' education, imputed cognition scores). For retrieving genetic information, I use the Polygenic Score Data. It contains polygenic scores (PGSs) for a variety of phenotypes for HRS respondents who provided salivary DNA between 2006 and 2012; from it, I retrieve the genetic scores for BMI and height.

Description & Transformation

Demographic Variables

Demographic variables include gender and year of birth, the sample contains individuals born between 1942 and 1980, with a prevalence of female (54.37%).

Table 1: Descriptive Statistics

	N	mean	sd	min	max
Gender	3,156	0.544	0.498	0	1
Year of Birth	3,156	1,954	5.565	1,942	1,980
Age (months)	3,156	679.3	66.27	368	829

Instrumental Variables and Exposure Variables

The BMIScore is a polygenetic score (PGS) created using results from a 2015 study conducted by the Genetic Investigation of ANthropometric Traits (GIANT) consortium. PGSs are obtained by computing the weighted sum of trait-associated alleles, in other words they combine different single nucleotide polymorphisms (SNPs). SNPs are the most common type of genetic variation in human DNA and each of them represents a difference in a single DNA nucleotide; in this case, all SNPs that influence BMI are included to compute the PGS. The PGS used here has been standardized within ethnicity to a standard normal curve (mean=0, standard deviation = 1) and refers to individuals of European ancestry. Similarly, the HeightScore is a PGS created by the same consortium in a study of 2014. Also in this case, the PGS was normalized within ethnicity. For what concerns the exposure variables, BMI is measured in kg/m^2 and presents an average of 28.68 kg/m^2 and a standard deviation of 5.897 kg/m^2 while height is measured in meters, with an average of 1.71 meters and a standard deviation of 10 cm. Descriptive statistics for the principal components can be found in Table A1.

Table 2: Descriptive Statistics

	N	mean	sd	min	max
BMIScore	3,156	0.00450	1.005	-3.633	3.637
HeightScore	3,156	-0.00882	0.985	-4.337	2.662
BMI	3,156	28.73	5.897	14.10	57.90
Height	3,156	1.709	0.100	1.410	2.083

Outcome Variables

YearsEduc is a continuous variable indicating the number of school years completed by an individual. HighEduc is a binary variable assuming a value of 1 if the individual has finished college and 0 otherwise. Income is a continuous variable and indicates individual earnings measured in US dollars; in order to simplify the interpretation and account for outliers, I take the natural logarithm. When taking the natural logarithm, observations with a value of 0 are not identified and are therefore dropped from the sample, with an elimination of about 400 observations. TotalWealth is a continuous variable and represents the net value of total wealth in US dollars, also in this case the logarithm is employed for the same reasons as before. Due to the presence of negative values and values of 0, also in this case there is a loss of about 300 observations. LaborStatus is a binary value taking a value of 1 if the individual is working full or part-time and a value of 0 if unemployed (retired and disabled individuals are excluded from the sample). Finally, Poverty is a binary variable that takes a value if the household of the individual has an income below the poverty threshold. The poverty threshold used by HRS is the one employed by the US Census Bureau and assumes different values based on the size of the family unit, the measurement I use here accounts for the size of the household already.

Table 3: Descriptive Statistics

	N	mean	sd	min	max
YearsEduc	3,137	14.09	2.197	0	17
TotalWealth	3,156	434,816	1.001e+06	-471,000	1.740e+07
Income	3,156	48,740	53,194	0	650,000
Poverty	3,156	0.0241	0.153	0	1
HighEduc	3,156	0.358	0.479	0	1
LogIncome	2,681	10.59	0.974	4.248	13.38
LogWealth	2,883	12.07	1.626	1.609	16.67
LaborStatus	3,156	0.918	0.274	0	1

Placebo Test Variables

Variables included in the placebo test follow the ones chosen by Böckerman et al. (2019). I employ mother's and father's education when available. Moreover, I also use different cognition scores (that could possibly be genes-related and affect SES). The NumberScore is created to follow a normal distribution with a mean of 500 and measures quantitative reasoning ability. The MentalScore comprehends scores from different tasks and measures mental ability as well as memory; the score ranges between 5 and 15.

Table 4: Descriptive Statistics

	N	mean	sd	min	max
MotherEduc	3,023	12.03	2.545	0	17
FatherEduc	2,875	11.81	3.230	0	17
NumberScore	3,085	522.1	30.58	390.2	579.6
MentalScore	1,577	13.59	1.532	5	15

Results

Placebo Tests

To provide robustness to the estimations, I perform a placebo test. Individuals are divided in two groups, one with below average BMI genetic score (LowBMIScore) and one with above average BMI genetic score (HighBMIScore). I compare the means of the two groups for a series of outcomes through a T-test to assess whether the two groups significantly differ (Table 5). Parental education is found to be significantly higher for people with a lower BMIscore. This finding could be reconducted to the presence of dynasty effects: parents with a lower BMIscore might achieve a higher education and then pass on their genes to their children. The independence assumption seems therefore to be a serious limitation in this analysis. To partially adjust for this, parental education can be included in the instrumental regression as a control variable. For what concerns the scores in mental tasks, I find no significant difference among the two groups for the general mental score, which provides support for the exclusion restriction, with the genes of BMI not having an effect on proxies of intelligence. On the contrary, a significantly higher numerical score is achieved for individuals with below average BMI. This could be suggestive of pleiotropy but, at the same time, it could be mediated by education (lower BMI might lead to higher achieved education which in turn might endow with better mathematical skills). As such, a definite conclusion on the matter of pleiotropy is not possible.

Table 5: Placebo Test: T-Test for Low and High BMIScores

	LowBMIScore	HighBMIScore	P-Value
FatherEduc	12.05688	11.56409	0.0000
MotherEduc	12.20606	11.85728	0.0002
NumberScore	523.3667	520.939	0.0275
MentalScore	13.62644	13.55764	0.3726

I complete the same placebo test dividing individuals according to their

genetic score for height (Table 6). In this case, the only significant difference between the two groups is found to be in the education of the mother. Also for height, therefore, a possible violation of the independence assumption seems to be present and the issue could be partially addressed by including parental education as a control variable in the instrumental regression. On the other side, the lack of difference in the mental capability scores is supportive of the exclusion restriction.

Table 6: Placebo Test: T-Test for Low and High HeightScores

	LowHeightScore	HighHeightScore	P-Value
FatherEduc	11.75021	11.84976	0.4152
MotherEduc	11.90087	12.12129	0.0187
NumberScore	521.1138	522.85	0.1197
MentalScore	13.56143	13.61574	0.4844

In order to further check for possible violations, I also perform a regression of the genetic scores (dependent variables) on the variables of interest, with the principal components and age as control variables (Table 7). These findings partially contradict the results of the T-test as only maternal education seem to have an impact and on the genetic score for height only. Nonetheless, this provides ground to the idea of adding parental education as a control variable in the upcoming regressions.

Table 7: Regression of BMIScore and HeightScore on Placebo Variables

	(1) BMIScore	(2) HeightScore
MotherEduc	-0.00527	0.0156**
	(0.0126)	(0.00707)
FatherEduc	-0.0120	-0.00184
	(0.00993)	(0.00558)
MentalScore	-0.0251	-0.00125
	(0.0183)	(0.0103)
NumberScore	-0.000342	-9.13e-05
	(0.000945)	(0.000531)
Constant	0.730	-0.0965
	(0.471)	(0.264)
Observations	1,393	1,393
R-squared	0.118	0.708
Controls	YES	YES

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

BMI and SES

The first stage regression for the pooled sample has an F-statistic of 19.05, this provides support to the first assumption of instrumental variables as the instrument, i.e. BMIscore, seems to be predictive of the exposure, i.e. BMI (Table A2). Moreover, the R^2 has a value of 10.8%, and is therefore in line with previous studies, which found a value of 6%; the difference may be due to the restriction of the analyzed sample (Yengo et al., 2018). Note, moreover, that the sample is uniformized across the different proxies, with a final count of 2.194 observations so to make the results comparable. For what concerns the effects on the outcome, I find significant negative effects of BMI in line with the expectations on three out of six proxies (Table 8).

For what concerns years of education, an increase of one unit of BMI (1 kg/m^2) leads to a decrease of 0.0692 years of education, ceteris paribus. Brumpton et al. (2019) found a similar yet more conservative estimate, with 1 additional unit of BMI decreasing years of education by 0.037. To make the results comparable to other studies, I include an interpretation in terms of standard deviations as well. One standard deviation (SD, 5.897 kg/m^2) increase in BMI leads to a decrease in education of 0.4 years, ceteris paribus. A SD increase in BMI also decreases by 0.015 the likelihood of completing higher education. For total wealth, an increase of one SD of BMI leads to a decrease in total wealth of 0.37%, all else being the same. No significant effects are found for the logarithm of income, for the employment status of the individuals and for the probability of being in poverty.

Table 8: Pooled Second Stage Regression for the Effect of BMI on SES

	(1)	(2)	(3)	(4)	(5)	(6)
	YearsEduc	HighEduc	LogIncome	LogWealth	LaborStatus	Poverty
BMI	-0.0692***	-0.0150***	-0.0188	-0.0622***	-0.00289	-0.00211*
	(0.0242)	(0.00565)	(0.0117)	(0.0181)	(0.00273)	(0.00123)
Constant	11.59***	-0.188	9.998***	10.23***	0.933***	0.0836*
	(0.956)	(0.223)	(0.462)	(0.716)	(0.108)	(0.0485)
Observations	2,194	2,194	2,194	2,194	2,194	2,194
R-squared	0.140	0.116	0.030	0.103	0.008	
Controls	YES	YES	YES	YES	YES	YES

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

To verify the existence of differences across genders, I rerun the regressions differentiating for men and women. The first stage regressions (Table A2) once more supports the assumption of relevance, with an F-statistic of 11.82 and 17.06 respectively. For men (Table 9), total wealth and higher education become insignificant at a 5% significance level. At the same time the coefficients of BMI on income and the likelihood of being in poverty, compared to the pool regression, become negative and significant. For women (Table 10), the coefficients of BMI for the two proxies of education become

insignificant at a 5% significance level. The only proxy that seems to be significantly affected by BMI for women is therefore total wealth. These results are in contrast with the findings of Cawley (2007) and Tyrrell et al. (2016), who found a causal effect of BMI on income for women only. The difference between genders may be due to the fact that different social standards may exist with respect to the weight of male and females and discrimination based on body size could be affected by this (Griffin, 2006).

Table 9: Second Stage Regression for the Effect of BMI on SES for Men Only

	(1)	(2)	(3)	(4)	(5)	(6)
	YearsEduc	HighEduc	LogIncome	LogWealth	LaborStatus	Poverty
BMI	-0.131***	-0.0201*	-0.0419**	-0.0583*	-0.00134	-0.00450**
	(0.0481)	(0.0109)	(0.0214)	(0.0351)	(0.00539)	(0.00203)
Constant	11.59***	-0.466	11.45***	9.528***	0.867***	0.180**
	(1.840)	(0.415)	(0.817)	(1.342)	(0.206)	(0.0774)
Observations	975	975	975	975	975	975
R-squared	0.109	0.112	0.030	0.085	0.014	
Controls	YES	YES	YES	YES	YES	YES

Standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 10: Second Stage Regression for the Effect of BMI on SES for Women Only

	(1) YearsEduc	(2) HighEduc	(3) LogIncome	(4) LogWealth	(5) LaborStatus	(6) Poverty
BMI	-0.0253	-0.0105*	-0.00544	-0.0605***	-0.00381	-0.000914
	(0.0268)	(0.00638)	(0.0133)	(0.0204)	(0.00304)	(0.00155)
Constant	11.49***	-0.0263	9.464***	10.54***	0.969***	0.0352
	(1.114)	(0.265)	(0.554)	(0.848)	(0.126)	(0.0645)
Observations	1,219	1,219	1,219	1,219	1,219	1,219
R-squared	0.166	0.130	0.032	0.131	0.009	0.006
Controls	YES	YES	YES	YES	YES	YES

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

All in all, causal effects are found for four indicators of SES, with differences between genders concerning income and education (effect only on men). The negative sign of the coefficients suggests that an increase in BMI has a negative causal effect on SES, as such the first hypothesis is not rejected.

Height and SES

The first stage regression for the pooled sample has an F-statistic of 21.15, this provides support to the first assumption of instrumental analysis (Table A3). The R^2 has a value of 9.6% that is lower than the 24% found by previous literature; by looking at men and women separately, however, the R^2 has a value of 18.7% and 21.5%, getting closer to the estimates of Yengo et al. (2018). For what concerns the effects on the outcome, significant positive effects of height in line with the expectations are found for only one out of six proxies of SES, namely income (Table 11). For income, an increase of one SD (10cm) in height leads to a decrease in income of 0.15%, ceteris paribus. As for BMI, no significant effects are found for the employment status of the individuals nor for the probability of being in poverty, moreover, no significant effects at a 5% significance level are present for the education

proxies and for total wealth. This seems to suggest that the effect of height on SES is less strong than the one for BMI. The motivation for this might be found in the mechanism: discrimination on the labor market, or in education, is more likely to be due to BMI than height, as BMI is a characteristic that seems more reflective of an individual's motivation and self-control.

Table 11: Pooled Second Stage Regression for the Effect of Height on SES

	(1)	(2)	(3)	(4)	(5)	(6)
	YearsEduc	HighEduc	LogIncome	LogWealth	LaborStatus	Poverty
Height	0.953	0.170	1.534**	2.101*	-0.0270	0.0957
	(1.567)	(0.366)	(0.747)	(1.201)	(0.178)	(0.0801)
Constant	7.733***	-0.963	6.845***	4.707**	0.882***	-0.144
	(2.636)	(0.615)	(1.256)	(2.020)	(0.300)	(0.135)
Observations	2,194	2,194	2,194	2,194	2,194	2,194
R-squared	0.152	0.125	0.072	0.074	0.005	
Controls	YES	YES	YES	YES	YES	YES

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

To verify the existence of differences across genders, I run the regressions again differentiating for men and women. The first stage regressions (Table A3) once more supports the assumption of relevance, with an F-statistic of 20.93 and 29.57 respectively. The estimation for each outcome can be found in Tables 12 and 13. For men, no significant effects are found for any of the proxies. For what concerns women, it is possible to notice a significant effect for income, while the other proxies are not significantly impacted by height. These results seem to contradict Tyrrell et al. (2016), who found a much stronger effect of height on income for men than for women. The results also differ from the conclusion of Hübler (2009), whose estimates for the effect of height on income were higher for men than women. On the contrary, the results seem to agree with the finding of Heineck (2008) whose effect of height on wages was more relevant for a female population.

21

Table 12: Second Stage Regression for the Effect of Height on SES for Men Only

	(1) YearsEduc	(2) HighEduc	(3) LogIncome	(4) LogWealth	(5) LaborStatus	(6) Poverty
Height	1.486	0.146	-0.538	2.632	-0.00207	0.144
	(2.688)	(0.617)	(1.228)	(2.019)	(0.310)	(0.114)
Constant	4.449	-1.410	11.03***	2.729	0.825	-0.237
	(5.018)	(1.151)	(2.292)	(3.770)	(0.578)	(0.213)
Observations	975	975	975	975	975	975
R-squared	0.159	0.133	0.031	0.084	0.011	
Controls	YES	YES	YES	YES	YES	YES

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 13: Second Stage Regression for the Effect of Height on SES for Women Only

	(1) YearsEduc	(2) HighEduc	(3) LogIncome	(4) LogWealth	(5) LaborStatus	(6) Poverty
Height	0.776	0.224	2.726***	1.935	-0.0817	0.0592
	(1.899)	(0.452)	(0.950)	(1.472)	(0.215)	(0.110)
Constant	9.349***	-0.754	4.753***	5.289**	0.976***	-0.0940
	(3.220)	(0.766)	(1.610)	(2.497)	(0.365)	(0.186)
Observations	1,219	1,219	1,219	1,219	1,219	1,219
R-squared	0.164	0.129	0.019	0.096	0.012	0.006
Controls	YES	YES	YES	YES	YES	YES

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

All in all, a causal effect of height is found only on income and seems to impact exclusively women. There is therefore a positive causal effect of height on at least one measure of SES, while for the other indicators no significant results are found. As such, there is no clear evidence for rejecting or not rejecting the second hypothesis.

Discussion & Conclusion

In this paper, I analyzed the causal effect of two physical appearance aspects, namely BMI and height, on different proxies of socio-economic status. To account for reverse causality and confounders, I employed the technique of Mendelian randomization, which uses genetic scores associated with BMI and height as instrumental variables. In doing so, I tried to isolate the causal effect that BMI and height have on different proxies of SES. The effect of BMI on SES was found to be negative and significant for years of education, the pursuit of higher education, income and total wealth, with differences between genders concerning income and education (effect only on men). The gender differences were in contrast with previous literature (Cawley, 2007; Tyrrell et al., 2016), which found an effect of BMI on income mostly for women. The reason for this discrepancy is not clear but it might concern different social standards in weight in the sample analyzed. On the other side, the effect of height was found to be positive and significant, with taller stature increasing income. This result, however, applied to women only. Moreover, no significant effects were found for the other proxies of SES. Altogether, these findings allow to address the research question:

"To what extent do body mass index (BMI) and height have a causal effect on socio-economic status (SES) indicators in a sample of American citizens?"

This analysis found evidence that physical appearance, in the form of height and especially BMI, plays an important role in determining different aspects of an individual SES, affecting education as well as labor market outcomes. In particular, higher BMI and shorter height, as estimated by genetics, causally lead to lower socio-economic status. These findings have important social and health implications, supporting evidence that overweight and shorter people are at disadvantage for what concerns many SES indicators. Policies could try to address the problem in different ways. The first solution would be to try to address the problem at its source and try to

reduce obesity. In this sense, policies could aim at increasing the availability and affordability of healthier food options while discouraging citizens to buy unhealthy food. The intervention could be implemented in different settings, from the promotion of healthy food in schools to an increase in taxation of fast-foods (Zhang, Liu, Liu, Xue, & Wang, 2014). Another solution would be to affect one of the possible mechanisms by which lower stature and higher BMI affect SES, namely the discrimination on the labor market. In this case, the government could intervene with laws and policies to prohibit discrimination. In this regard, Puhl et al. (2015) found evidence that public support in the United States, Canada, and Australia agrees that the government should have specific laws in place to prohibit weight discrimination. Similarly, jurisdiction could protect from height discrimination. An example that could be followed is the one of Victoria in Australia, which included height in its prohibition on discrimination based on physical appearance under the Equal Opportunity Act of 1995 (Charlesworth, 2008).

Despite the results being in line with the predictions and with past research, this analysis is still subject to several limitations in terms of both internal and external validity. For what concerns internal validity, although the technique of Mendelian randomization is helpful in alleviating issues of confounding and reverse causality, problems associated with using genes as instrumental variables are still present. As already introduced in the methodology, the possible violation of the independence assumption is a particularly severe problem. In this case, dynasty effects could be an issue with parental genes having an effect both on the genes of the offspring and the SES of the offspring during childhood. I tried to partially address this issue by controlling for parental education, there might still be, however, a bias in the estimation of the effect and an overestimation of the coefficients. In this regard, Brumpton et al. (2019) found that, after controlling for family fixed effects by mean of a sibling study, the effects of BMI and height on education largely decreased in within-family analyses. Another limitation derives from the variables employed. First, BMI is not a perfect approximation for obesity. As Johansson, Böckerman, Kiiskinen, and Heliövaara (2009) argue, BMI

24

does not take into consideration the proportion of fat mass of an individual; other measurements, like waist circumference, would be more suitable for the purpose. Secondly, the proxies for SES may not cover the entirety of social status and may be subject to measurement error as they are self-reported. Another limitation of this research concerns the unavailability of BMI and height at the time that education was completed. This leads to possibly biased results and eliminates the possibility of accounting for growth sprouts, which were found to be an important factor for the effect of height on SES by past literature (Persico et al., 2004). For what concerns the external validity, only people of European descent are taken into consideration due to data availability. This prevents from extending the conclusions to individuals of African descendent that might be differently affected by BMI and stature. The sample comprehended, moreover, individuals aged between 30 and 68 and therefore focuses on an older population, not considering people in their twenties. These individuals might be impacted by BMI and height differently, Norton and Han (2008), for example, found little to no effect for this category.

Despite its limitations, this paper can be the starting point of future research. The first possibility would be to reproduce the results while trying to solve the violation of the independence assumption. To do so, a more comprehensive dataset would be needed, containing either information about parental genes or about siblings. Controlling for parents' genes would allow to mitigate for dynasty effects, the same results could be obtained by using family fixed effects, on the model of Brumpton et al. (2019). Another suggestion for further research could be the analysis of the studied phenomenon for individuals at different stages of their lives. This would allow to check whether appearance has a different effect on SES at the start or at later stages of an individual's working life. Finally, research might aim at exploring the mechanism by which lower stature and higher BMI lead to lower SES. As discussed, one of the possible mechanisms is the discrimination on the labor market. To assess whether such discrimination is actually taking place, a design similar to the one of Bertrand and Mullainathan (2004) could be implemented. Fictitious resumes could be send, presenting people with

the same qualifications but with pictures indicative of a different BMI. A significant difference in callbacks rates would be suggestive of discrimination based on BMI and provide support for the aforementioned mechanism.

References

Adler, N. E., Boyce, T., Chesney, M. A., Cohen, S., Folkman, S., Kahn, R. L., & Syme, S. L. (1994). Socioeconomic status and health: The challenge of the gradient. *American psychologist*, *49*(1), 15.

Atella, V., Pace, N., & Vuri, D. (2008). Are employers discriminating with respect to weight?: European evidence using quantile regression. *Economics & Human Biology*, *6*(3), 305–329.

Bann, D., Johnson, W., Li, L., Kuh, D., & Hardy, R. (2017). Socioeconomic inequalities in body mass index across adulthood: Coordinated analyses of individual participant data from three british birth cohort studies initiated in 1946, 1958 and 1970. *PLoS medicine*, *14*(1), e1002214.

Bertrand, M., & Mullainathan, S. (2004). Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. *American economic review*, *94*(4), 991–1013.

Böckerman, P., Cawley, J., Viinikainen, J., Lehtimäki, T., Rovio, S., Seppälä, I., . . . Raitakari, O. (2019). The effect of weight on labor market outcomes: An application of genetic instrumental variables. *Health economics*, *28*(1), 65–77.

Brumpton, B., Sanderson, E., Hartwig, F. P., Harrison, S., Vie, G. A., Cho, Y., . . . Havdahl, A., et al. (2019). Within-family studies for mendelian randomization: Avoiding dynastic, assortative mating, and population stratification biases. *bioRxiv*, 602516.

Brunello, G., & d'Hombres, B. (2007). Does body weight affect wages?: Evidence from europe. *Economics & Human Biology*, *5*(1), 1–19.

Cawley, J. (2007). The labor market impact of obesity. *Obesity, business and public policy*, 76–86.

Charlesworth, S. (2008). Claiming discrimination, complaints of sex and gender discrimination in employment under the victorian equal opportunity act 1995. *School of Global Studies, Social Science & Planning Working Paper Series*, (1).

Cinnirella, F., & Winter, J. K. (2009). Size matters! body height and labor market discrimination: A cross-european analysis. *SSRN Electronic Journal*.

Conley, D., & Glauber, R. (2006). Gender, body mass, and socioeconomic status: New evidence from the psid. In *The economics of obesity* (pp. 253–275). Emerald Group Publishing Limited.

Frieze, I. H., Olson, J. E., & Good, D. C. (1990). Perceived and actual discrimination in the salaries of male and female managers 1. *Journal of Applied Social Psychology*, *20*(1), 46–67.

Greve, J. (2008). Obesity and labor market outcomes in denmark. *Economics & Human Biology*, *6*(3), 350–362.

Griffin, A. W. (2006). Women and weight-based employment discrimination. *Cardozo JL & Gender*, *13*, 631.

Hartwig, F. P., Davies, N. M., & Davey Smith, G. (2018). Bias in mendelian randomization due to assortative mating. *Genetic epidemiology*, *42*(7), 608–620.

Heineck, G. (2008). A note on the height–wage differential in the uk–cross-sectional evidence from the bhps. *Economics Letters*, *98*(3), 288–293.

Hensley, W. E. (1993). Height as a measure of success in academe. *Psychology: a journal of human behavior*.

Hu, F. (2008). *Obesity epidemiology*. Oxford University Press.

Hübler, O. (2009). The nonlinear link between height and wages in germany, 1985–2004. *Economics & Human Biology*, *7*(2), 191–199.

Jansen, W., & Hazebroek-Kampschreur, A. (1997). Differences in height and weight between children living in neighbourhoods of different socioeconomic status. *Acta Paediatrica*, *86*(2), 224–225.

Johansson, E., Böckerman, P., Kiiskinen, U., & Heliövaara, M. (2009). Obesity and labour market success in finland: The difference between having a high bmi and being fat. *Economics & Human Biology*, *7*(1), 36–45.

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*(3), 428.

Kim, T. J., & von dem Knesebeck, O. (2018). Income and obesity: What is the direction of the relationship? a systematic review and meta-analysis. *BMJ open, 8*(1), e019862.

Lin, C. C., Rogot, E., Johnson, N. J., Sorlie, P. D., & Arias, E. (2003). A further study of life expectancy by socioeconomic factors in the national longitudinal mortality study. *Ethnicity & disease, 13*(2), 240–247.

Magnusson, P. K., Rasmussen, F., & Gyllensten, U. B. (2006). Height at age 18 years is a strong predictor of attained education later in life: Cohort study of over 950 000 swedish men. *International Journal of Epidemiology, 35*(3), 658–663.

Marmot, M. (2017). The health gap: The challenge of an unequal world: The argument. *International journal of epidemiology, 46*(4), 1312–1318.

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. *Jama, 272*(11), 859–866.

Morris, S. (2007). The impact of obesity on employment. *Labour Economics, 14*(3), 413–433.

Murasko, J. E. (2009). Socioeconomic status, height, and obesity in children. *Economics & Human Biology, 7*(3), 376–386.

Norton, E. C., & Han, E. (2008). Genetic information, obesity, and labor market outcomes. *Health economics, 17*(9), 1089–1104.

Persico, N., Postlewaite, A., & Silverman, D. (2004). The effect of adolescent experience on labor market outcomes: The case of height. *Journal of Political Economy, 112*(5), 1019–1053.

Puhl, R. M., & Latner, J. D. (2007). Stigma, obesity, and the health of the nation's children. *Psychological bulletin, 133*(4), 557.

Puhl, R. M., Latner, J. D., O'brien, K. S., Luedicke, J., Danielsdottir, S., & Salas, X. R. (2015). Potential policies and laws to prohibit weight discrimination: Public views from 4 countries. *The Milbank Quarterly, 93*(4), 691–731.

Rooth, D.-O. (2007). Evidence of unequal treatment in hiring against obese applicants: A field experiment. *SSRN Electronic Journal.*

Rooth, D.-O. (2009). Obesity, attractiveness, and differential treatment in hiring a field experiment. *Journal of human resources*, *44*(3), 710–735.

Rosenberg, I. B. (2009). Height discrimination in employment. *Utah L. Rev.* 907.

Roszell, P., Kennedy, D., & Grabb, E. (1989). Physical attractiveness and income attainment among canadians. *The journal of psychology*, *123*(6), 547–559.

Smith, J. P. (2007). The impact of socioeconomic status on health over the life-course. *Journal of Human Resources*, *42*(4), 739–764.

Sobal, J., & Stunkard, A. J. (1989). Socioeconomic status and obesity: A review of the literature. *Psychological bulletin*, *105*(2), 260.

Speliotes, E. K., Willer, C. J., Berndt, S. I., Monda, K. L., Thorleifsson, G., Jackson, A. U., ... Mägi, R., et al. (2010). Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature genetics*, *42*(11), 937.

Tyrrell, J., Jones, S. E., Beaumont, R., Astley, C. M., Lovell, R., Yaghootkar, H., ... Hirschhorn, J. N., et al. (2016). Height, body mass index, and socioeconomic status: Mendelian randomisation study in uk biobank. *bmj*, *352*, i582.

Verbanck, M., Chen, C.-Y., Neale, B., & Do, R. (2018). Detection of widespread horizontal pleiotropy in causal relationships inferred from mendelian randomization between complex traits and diseases. *Nature genetics*, *50*(5), 693.

Vyas, S., & Kumaranayake, L. (2006). Constructing socio-economic status indices: How to use principal components analysis. *Health policy and planning*, *21*(6), 459–468.

Yengo, L., Sidorenko, J., Kemper, K. E., Zheng, Z., Wood, A. R., Weedon, M. N., ... Visscher, P. M., et al. (2018). Meta-analysis of genome-wide association studies for height and body mass index in 700000 individuals of european ancestry. *Human molecular genetics*, *27*(20), 3641–3649.

Young, T. J., & French, L. A. (1996). Height and perceived competence of us presidents. *Perceptual and Motor Skills*, *82*(3), 1002–1002.

Zhang, Q., Liu, S., Liu, R., Xue, H., & Wang, Y. (2014). Food policy approaches to obesity prevention: An international perspective. *Current obesity reports*, *3*(2), 171–182.

Appendices

Table A1: Descriptive Statistics

	N	mean	sd	min	max
PC1_5A	3,156	5.91e-05	0.00875	-0.0357	0.0447
PC1_5B	3,156	-0.000305	0.00910	-0.0507	0.0128
PC1_5C	3,156	-0.000273	0.00907	-0.0234	0.0198
PC1_5D	3,156	-0.000275	0.00909	-0.0410	0.0319
PC1_5E	3,156	-0.000150	0.00887	-0.0427	0.0110
PC6_10A	3,156	-0.000127	0.00894	-0.0317	0.0348
PC6_10B	3,156	-1.12e-05	0.00882	-0.0326	0.0294
PC6_10C	3,156	0.000124	0.00926	-0.0342	0.0335
PC6_10D	3,156	-1.75e-05	0.00900	-0.0291	0.0346
PC6_10E	3,156	0.000155	0.00894	-0.0402	0.0261

Table A2: First Stage Regression for the Effect of BMIScore on BMI

| | (1) | (2) | (3) |
	BMI	BMI (Men Only)	BMI (Women Only)
BMIScore	1.907***	1.564***	2.199***
	(0.118)	(0.154)	(0.172)
Constant	29.79***	29.36***	30.94***
	(1.191)	(1.697)	(1.643)
Controls	YES	YES	YES
Observations	2,458	1,095	1,363
R-squared	0.106	0.095	0.133
F-Stat	19.05	11.82	17.06

Standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A3: First Stage Regression for the Effect of HeightScore on Height

	(1)	(2)	(3)
	Height	Height (Men Only)	Height (Women Only)
HeightScore	0.0483***	0.0460***	0.0525***
	(0.00357)	(0.00356)	(0.00298)
Constant	1.704***	1.864***	1.700***
	(0.0203)	(0.0215)	(0.0163)
Controls	YES	YES	YES
Observations	2,458	1,095	1,363
R-squared	0.081	0.167	0.215
F-stat	21.15	20.93	29.57

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1