Height, body mass index, and socioeconomic status: mendelian randomisation study in UK Biobank

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.
Height, body mass index, and socioeconomic status: mendelian randomisation study in UK Biobank

Jessica Tyrrell,1,2 Samuel E Jones,1 Robin Beaumont,1 Christina M Astley,3,4 Rebecca Lovell,2 Hanieh Yaghootkar,1 Marcus Tukey,1 Katherine S Ruth,1 Rachel M Freathy,1 Joel N Hirschhorn,2,3,5 Andrew R Wood,1 Anna Murray,1 Michael N Weedon,1 Timothy M Frayling1

OBJECTIVE
To determine whether height and body mass index (BMI) have a causal role in five measures of socioeconomic status.

DESIGN
Mendelian randomisation study to test for causal effects of differences in stature and BMI on five measures of socioeconomic status. Mendelian randomisation exploits the fact that genotypes are randomly assigned at conception and thus not confounded by non-genetic factors.

SETTING
UK Biobank.

PARTICIPANTS
119 669 men and women of British ancestry, aged between 37 and 73 years.

MAIN OUTCOME MEASURES
Age completed full time education, degree level education, job class, annual household income, and Townsend deprivation index.

RESULTS
In the UK Biobank study, shorter stature and higher BMI were observationally associated with several measures of lower socioeconomic status. The associations between shorter stature and lower socioeconomic status tended to be stronger in men, and the associations between higher BMI and lower socioeconomic status tended to be stronger in women. For example, a 1 standard deviation (SD) higher BMI was associated with a £210 (£276; $300; 95% confidence interval £84 to £420; P=6×10−7) lower annual household income in men and a £1890 (£1680 to £2100; P=6×10−15) lower annual household income in women. Genetic analysis provided evidence that these associations were partly causal. A genetically determined 1 SD (6.3 cm) taller stature caused a 0.06 (0.02 to 0.09) year older age of completing full time education (P=0.01), a 1.12 (1.07 to 1.18) times higher annual household income (P=4×10−7), and a £1130 (£680 to £1580) higher annual household income and a 0.10 (0.04 to 0.16) SD (P=0.001) higher level of deprivation in women only.

CONCLUSIONS
These data support evidence that height and BMI play an important partial role in determining several aspects of a person’s socioeconomic status, especially women’s BMI for income and deprivation and men’s height for education, income, and job class. These findings have important social and health implications, supporting evidence that overweight people, especially women, are at a disadvantage and that taller people, especially men, are at an advantage.

Introduction
Higher socioeconomic status is associated with better health and longer life.12 For example, a recent article highlighted the strength of the association between income and lower BMI and men’s height for education, income, and job class. These findings have important social and health implications, supporting evidence that overweight people, especially women, are at a disadvantage and that taller people, especially men, are at an advantage.

Socioeconomic status influences morbidity and mortality, with a recent review highlighting the 18 year gap in life expectancy between men living in the poorest and richest boroughs of London.3
Taller stature and lower body mass index (BMI) are associated with higher socioeconomic status, but the causal directions of these associations are poorly understood.4
Understanding the causal directions of these associations is important for public health and wellbeing policies.5

This study provides a high level of evidence (using 119 000 participants from the UK Biobank) for a causal effect from shorter stature and higher BMI to lower measures of socioeconomic status.5
Shorter height, as estimated by genetics, leads to lower levels of education, lower job status, and less income in men in particular, and higher BMI leads to lower income and greater deprivation in women.6
Genetic evidence has the advantage of being largely free from the problems that afflict observational studies; analyses using inherited DNA variation are much more robust to confounding, bias, and reverse causality.7

Evidence as to whether height and BMI have causal effects on socioeconomic status through these, or other, pathways is limited. For example, to our knowledge, no large studies have compared siblings or twins of...
different heights and BMIs, where childhood environment could be controlled for. If differences in BMI and height can lead to differences in socioeconomic status, this would have implications for policy makers. For example, evidence of a causal link would further highlight the need to adjust for unconscious biases in decision making in education and employment.

Gene based analyses, such as mendelian randomisation, can be used to test for a causal relation between socioeconomic status and a genetically influenced phenotype such as BMI. Genetic variants can act as unconfounded proxies for the risk factors under investigation—here, BMI and height—because inherited genetic variation is randomly allocated at conception. The outcomes being tested—here, measures of socioeconomic status—cannot influence genetic variation, so reverse causality is avoided in genetic studies. Figure 1 illustrates the principle of mendelian randomisation. Previous studies have used genetic variants to test causal relation between health traits such as BMI and socioeconomic status related outcomes such as academic performance. However, these studies were limited by a lack of genetic variants robustly associated with BMI and by sample sizes of fewer than 2300 people. Recent genome-wide association studies have identified many 10s and 100s of genetic variants associated with BMI and height, respectively, and so provide the tools for mendelian randomisation tests.

Here, we used mendelian randomisation analysis to test the hypothesis that causal pathways link BMI and height to differences in five different measures of socioeconomic status. We used the first release of data from the UK Biobank. The UK Biobank has 119 669 participants of white British ancestry with genetic data, measures of socioeconomic status, and height and BMI measures. The UK Biobank thus represents a very powerful resource in which to investigate the causal relation between BMI, height, and socioeconomic status by using mendelian randomisation analysis.

Methods

UK Biobank

The UK Biobank recruited more than 500 000 people aged 37-73 years (99.5% were between 40 and 69 years) from across the country in 2006-10. Participants provided a range of information via questionnaires and interviews (such as demographics, health status, and lifestyle); anthropometric measurements, blood pressure readings, and blood, urine and saliva samples were taken for future analysis. This has been described in more detail elsewhere. We used 120 286 participants of white British descent from the initial UK Biobank dataset, of whom 119 669 had valid genetic data and both BMI and height measures available. We did not include other ethnic groups, because individually they were underpowered. Table 1 shows the basic characteristics of the sample. We defined people of white British descent as those who both self identified as white British and were confirmed as ancestrally “Caucasian” using principal components analyses of genome-wide genetic information. This dataset underwent extensive central quality control (http://biobank.ctsu.ox.ac.uk) (see supplementary methods).

Patient involvement

This study was conducted using the UK Biobank resource. Details of patient and public involvement in the UK Biobank are available online (www.ukbiobank.ac.uk/about-biobank-uk/) and https://www.ukbiobank.ac.uk/wp-content/uploads/2011/07/Summary-EGF-consultation.pdf?phpMyAdmin=trmKQjYdjnnQIgL%2CfAzikMh%3Ex). No patients were specifically involved in setting the research question or the outcome measures, nor were they involved in developing plans for recruitment, design, or implementation of this study. No patients were asked to advise on interpretation or writing up of results. There are no specific plans to disseminate the results of the research to study participants, but the UK Biobank disseminates key findings from projects on its website.

Exposure and outcome measures

Exposure and outcome measures were all collected at baseline when participants attended the assessment centre.

Height

Height (cm) was measured using a Seca 202 device in all participants in the UK Biobank (n=500 120). Sitting height was also measured (n=496 380). We excluded one person with a height more than 4.56 SD away from the mean and a sitting height to standing height ratio of greater than 0.75, which is not compatible with normal growth. In all, 119 669 people of white British ancestry with genetic data available also had a valid height and BMI measure.
Table 1 | Summary of demographic characteristics of 119 669 participants of white British ancestry with valid genetic data and height and body mass index measures available, stratified by sex. Values are numbers (percentages) unless stated otherwise

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>All (n=119 669)</th>
<th>Men (n=56 652)</th>
<th>Women (n=63 017)</th>
<th>P value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (SD) age at recruitment, years</td>
<td>56.9 (7.9)</td>
<td>57.3 (8.0)</td>
<td>56.6 (7.8)</td>
<td>&lt;1x10^{-15}</td>
</tr>
<tr>
<td>Male sex</td>
<td>56 652 (47.3)</td>
<td>56 652 (100)</td>
<td>63 017 (100)</td>
<td>–</td>
</tr>
<tr>
<td>Mean (SD) body mass index, kg/m²</td>
<td>27.5 (4.8)</td>
<td>27.9 (4.3)</td>
<td>27.2 (5.2)</td>
<td>&lt;1x10^{-15}</td>
</tr>
<tr>
<td>Mean (SD) height, cm</td>
<td>168.8 (9.2)</td>
<td>175.7 (6.7)</td>
<td>162.6 (6.2)</td>
<td>&lt;1x10^{-15}</td>
</tr>
<tr>
<td>Smoking status:</td>
<td></td>
<td></td>
<td></td>
<td>&lt;1x10^{-15}</td>
</tr>
<tr>
<td>Never</td>
<td>63 806 (53.3)</td>
<td>27 834 (49.1)</td>
<td>35 972 (57.1)</td>
<td>–</td>
</tr>
<tr>
<td>Former</td>
<td>40 890 (34.2)</td>
<td>21 162 (37.4)</td>
<td>19 728 (31.3)</td>
<td>–</td>
</tr>
<tr>
<td>Current</td>
<td>13 332 (11.1)</td>
<td>6767 (11.9)</td>
<td>6565 (10.4)</td>
<td>–</td>
</tr>
<tr>
<td>Missing</td>
<td>1641 (1.4)</td>
<td>889 (1.6)</td>
<td>752 (1.2)</td>
<td>–</td>
</tr>
<tr>
<td>Mean (SD) age completed full time education, years</td>
<td>16.6 (2.2)</td>
<td>16.6 (2.4)</td>
<td>16.5 (2.2)</td>
<td>2x10^{-7}</td>
</tr>
<tr>
<td>Degree level education</td>
<td>53 652 (44.8)</td>
<td>25 956 (45.8)</td>
<td>27 696 (44.0)</td>
<td>6x10^{-15}</td>
</tr>
<tr>
<td>Job class:</td>
<td></td>
<td></td>
<td></td>
<td>&lt;1x10^{-15}</td>
</tr>
<tr>
<td>Elementary occupations</td>
<td>3912 (3.3)</td>
<td>2054 (3.6)</td>
<td>1878 (3.0)</td>
<td>–</td>
</tr>
<tr>
<td>Process plant and machine operatives</td>
<td>3760 (3.1)</td>
<td>1821 (3.2)</td>
<td>1939 (3.1)</td>
<td>–</td>
</tr>
<tr>
<td>Sales and customer service occupations</td>
<td>2658 (2.2)</td>
<td>1321 (2.3)</td>
<td>1337 (2.1)</td>
<td>–</td>
</tr>
<tr>
<td>Leisure and other personal service occupations</td>
<td>963 (0.8)</td>
<td>452 (0.8)</td>
<td>511 (0.8)</td>
<td>–</td>
</tr>
<tr>
<td>Personal service occupations</td>
<td>3567 (3.0)</td>
<td>1693 (2.9)</td>
<td>1874 (3.0)</td>
<td>–</td>
</tr>
<tr>
<td>Skilled trades</td>
<td>1907 (1.6)</td>
<td>900 (1.6)</td>
<td>1007 (1.6)</td>
<td>–</td>
</tr>
<tr>
<td>Administrative and secretarial roles</td>
<td>11878 (9.9)</td>
<td>5759 (10.3)</td>
<td>6119 (9.8)</td>
<td>–</td>
</tr>
<tr>
<td>Business and public sector associate professionals</td>
<td>4631 (3.9)</td>
<td>2548 (4.5)</td>
<td>2083 (3.3)</td>
<td>–</td>
</tr>
<tr>
<td>Associate professionals</td>
<td>8388 (7.0)</td>
<td>4219 (7.5)</td>
<td>4169 (6.6)</td>
<td>–</td>
</tr>
<tr>
<td>Professional occupations</td>
<td>17044 (14.2)</td>
<td>8934 (15.8)</td>
<td>8110 (12.9)</td>
<td>–</td>
</tr>
<tr>
<td>Senior officials</td>
<td>13526 (11.3)</td>
<td>7424 (13.2)</td>
<td>6102 (9.8)</td>
<td>–</td>
</tr>
<tr>
<td>Income:</td>
<td></td>
<td></td>
<td></td>
<td>&lt;1x10^{-15}</td>
</tr>
<tr>
<td>&lt;£18 000</td>
<td>23817 (19.9)</td>
<td>12600 (22.6)</td>
<td>11217 (17.7)</td>
<td>–</td>
</tr>
<tr>
<td>£18 000 to £30 999</td>
<td>26808 (22.4)</td>
<td>14288 (25.2)</td>
<td>12520 (19.8)</td>
<td>–</td>
</tr>
<tr>
<td>£31 000 to £51 999</td>
<td>27245 (22.8)</td>
<td>14388 (25.4)</td>
<td>12857 (19.8)</td>
<td>–</td>
</tr>
<tr>
<td>£52 000 to £100 000</td>
<td>20397 (17.0)</td>
<td>10950 (19.3)</td>
<td>9447 (15.0)</td>
<td>–</td>
</tr>
<tr>
<td>&gt;£100 000</td>
<td>5060 (4.2)</td>
<td>2777 (4.9)</td>
<td>2283 (3.6)</td>
<td>–</td>
</tr>
<tr>
<td>Mean (SD) Townsend deprivation index</td>
<td>−1.5 (3.0)</td>
<td>−1.5 (3.0)</td>
<td>−1.5 (3.0)</td>
<td>–</td>
</tr>
<tr>
<td>Overall per allele height SNP association with height</td>
<td>0.022 (0.021 to 0.023); P&lt;1x10^{-15}</td>
<td>0.022 (0.021 to 0.023); P&lt;1x10^{-15}</td>
<td>0.020 (0.020 to 0.021); P&lt;1x10^{-15}</td>
<td>–</td>
</tr>
<tr>
<td>Overall per allele BMI SNP association with BMI</td>
<td>0.022 (0.021 to 0.023); P&lt;1x10^{-15}</td>
<td>0.022 (0.021 to 0.023); P&lt;1x10^{-15}</td>
<td>0.025 (0.023 to 0.026); P&lt;1x10^{-15}</td>
<td>–</td>
</tr>
</tbody>
</table>

BMI=body mass index; SNP=single nucleotide polymorphism.
Not all five socioeconomic status measures were available in all 119 669 individuals (see supplementary table A for further information).
*For comparison between men and women; models were adjusted for age at recruitment.

Body mass index
The UK Biobank has two different measures of BMI—one calculated as weight/height² and one measured using electrical impedance to quantify mass. We excluded people with significant differences (>4.56 SD from the mean) between impedance and normal BMI measures (n=1172) where both variables were available. If only one measure of BMI was available, we used this (n=7290). Valid BMI was available for 119 669 people with genetic and height data available.

Socioeconomic status
We used five different socioeconomic status variables.¹ Age in years at completion of full time education (questionnaire based; available only for people who did not go on to degree level education). Data were available for 82 543 people and missing in 37 126 people with valid height, BMI, and genetic data.² Education (coded as degree level or not, derived from the questionnaire); participants were asked, “Which of the following qualifications do you have? (You can select more than one),” with the options college or university degree, A levels or equivalent, O levels or GCSEs or equivalent, CSEs, NVQ/HND/ HNC, professional qualifications (eg, nursing or teaching). We created a dichotomous variable comparing degree level education or professional qualifications (n=53652) with other qualifications (n=64913); 110 people did not respond to this question.³ Job class (coded as elementary occupations, process plant and machine operatives, sales and customer service occupations, leisure and other personal service occupations, personal service occupations, skilled trades, administrative and secretarial roles, business and public sector associate professionals, associate professionals, professional occupations, and managers and senior officials); this was coded from the UK Biobank job code variable. All participants were asked to select their current or most recent job. Data were available for 76 404 people, with missing data in 43265. We dichotomised this variable into unskilled (n=21036); elementary occupations to personal service occupations, and managers and senior officials; 21 036; elementary occupations to personal service occupations, and managers and senior officials; this was coded from the UK Biobank job code variable. All participants were asked to select their current or most recent job. Data were available for 76 404 people, with missing data in 43265. We dichotomised this variable into unskilled (n=21036); elementary occupations to personal service occupations, and managers and senior officials; 21 036; elementary occupations to personal service occupations, and managers and senior officials; this was coded from the UK Biobank job code variable. All participants were asked to select their current or most recent job.
and missing in 16 432 people with valid height, BMI, and genetic data. Townsend deprivation index (a composite measure of deprivation based on unemployment, non-car ownership, non-home ownership, and household overcrowding; a negative value represents high socioeconomic status). This was calculated before participants joined the UK Biobank and was based on the preceding national census data, with each participant assigned a score corresponding to the postcode of their home dwelling. Data were available for 119 519 people and missing in 150 people with valid height, BMI, and genetic data. For each of the five traits, we compared participants missing data with those reporting data; generally, those with missing data were older and shorter with higher BMIs (supplementary table A). We investigated the relation of these five socioeconomic status measures and four health outcomes: self reported coronary artery disease, hypertension (defined as a systolic blood pressure ≥140, a diastolic blood pressure ≥90, or the report of use of blood pressure lowering drugs), any self reported long term illness (based on the UK Biobank question “Do you have a long standing illness, disability or infirmity?”), and type 2 diabetes (based on self report, excluding people using insulin in the first year after diagnosis and those given a diagnosis before 35 years of age or within the previous year) (supplementary table B). For three of the traits (the exceptions being education and job class, both binary traits), we converted the data to a normal distribution to limit the influence of any subtle population stratification and to provide standard deviation effect sizes. We took residuals of the exposure and outcome measures from standard linear regression by using nine covariates: age, sex, assessment centre location, five (within UK) ancestry principal components, and microarray used to measure genotypes. We then inverse normalised the residualised variables. To convert our results back to meaningful units after mendelian randomisation, we multiplied our SD $\beta$s by a 1 SD change in the socioeconomic status measure. For example, a 1 SD change in Townsend deprivation index was equivalent to 2.68 units. Therefore, a 0.05 SD change equated to a 0.134 unit change in deprivation.

Observational associations
We regressed each socioeconomic status measure against height and BMI by using linear regression for continuous outcome variables and logistic regression for binary outcomes. We adjusted these associations for age and sex. We also investigated the association of each socioeconomic status measure with a range of health outcomes.

Genetic variants
The genetic variants used were extracted genotypes from UK Biobank’s imputation dataset (the supplementary methods provide more information on the UK Biobank’s quality control). We excluded individual genotypes if the genotype probability was less than 0.9. We confirmed that the variants were imputed with high quality by comparing them with the directly genotyped data, where available. Details of imputation quality are given in supplementary table C.

Height—We selected 396 of 404 height genetic variants from independent loci that were associated with height at genome-wide significance in the GIANT studies of up to 253 288 people (supplementary table C). We excluded eight variants that were unavailable (rs1420023, rs567401), were poorly imputed, had an imputation quality <0.9 (rs11683207, rs7534365), or were not in Hardy-Weinberg equilibrium (P<1×10−5; rs1401795, rs7692995, rs915506, rs3790086). The 396 variants explained 12.3% of the variance in adult height in the UK Biobank participants used.

Body mass index—We selected 69 of 76 common genetic variants that were associated with BMI at genome-wide significance in the GIANT consortium in studies of up to 339 224 people (supplementary table C). We limited the BMI variants to those that were associated with BMI in the analysis of all people of European ancestry and did not include those that reached genome-wide levels of statistical confidence in only one sex or one stratum. We also excluded variants if they were known to be classified as a secondary signal within a locus. Three variants were excluded from the score owing to potential pleiotropy (rs11030104 (BDNF reward phenotypes), rs13107325 (SLC39A8 lipids, blood pressure), rs3888190 (SH2B1 multiple traits)), three were not in Hardy-Weinberg equilibrium (P<1×10−6; rs17001654, rs2075650, rs9925964), and one was unavailable (rs2033529). The 69 variants explained 1.5% of the variance in BMI in the UK Biobank participants.

We recoded individual variants as 0, 1, and 2 according to the number of height or BMI increasing alleles. We used the variants to create height and BMI genetic risk scores. Each variant was weighted by its relative effect size ($\beta$ coefficient) obtained from the reported meta-analysis data. We created a weighted score, in which $\beta$ is the $\beta$ coefficient of representing the association between each single nucleotide polymorphism (SNP) and height/BMI: weighted score=$\sum_{i=1}^{n} \beta_i \times SNP_i$; where $\sum_{i=1}^{n} \beta_i \times SNP_i$. We then rescaled the weighted score to reflect the number of trait increasing alleles: weighted genetic risk score=(weighted score/number of SNPs)/sum of $\beta$ coefficients.

Mendelian randomisation
The mendelian randomisation approach used in this study made the following assumptions: the height and BMI genetic risk scores were robustly associated with measured height and BMI; the height and BMI genetic risk scores were not associated with confounding factors that bias conventional epidemiological associations between height/BMI and socioeconomic status; the height and BMI genetic risk scores were related to the outcome only via its association with the modifiable exposure; and the associations represented in figure 1 are linear and unaffected by statistical interactions.

Instrumental variable analysis
We used two methods that use genetic variants to assess causal relations between two traits. Firstly, to estimate the causal effect of height or BMI on individual socioeconomic status measures, we used instrumental
variable analysis using the height or BMI genetic risk score. We used the two stage, least squares estimator method that uses predicted levels of BMI or height per genotype and regresses each outcome against these predicted values.

For continuous socioeconomic status outcomes, we used the ivreg2 command in Stata to do the instrumental variable analysis. We compared results from observational and instrumental variable regressions by using the Durbin-Wu-Hausman test for endogeneity, which examines the difference between the estimates from linear regression (observational) and instrumental variable analysis.

For binary outcomes, the instrumental variable analysis was done in two stages. Firstly, we assessed the association between the height or BMI genetic risk score and height or BMI, respectively. We saved the predicted values and residuals from this regression model. Secondly, we used the predicted values from stage 1 as the independent variable (reflecting an unconfounded estimate of variation in BMI or height) and degree status or job class as the dependent variable in a logistic or ordinal logistic regression model. We used robust standard errors to correct for uncertainty in the estimate. We examined the F statistics from first stage regressions to evaluate the strength of the instruments; weak instruments can bias results towards the (confounded) multivariable regression association or towards the null in a two stage design.

Egger method
We used a second method of mendelian randomisation, the Egger method, as a sensitivity analysis if the instrumental variables test result was noteworthy. This method is more robust to potential violations of the standard instrumental variable assumptions. It uses a weighted regression with an unconstrained intercept to regress the effect sizes of variant-outcome associations weighted regression with an unconstrained intercept to standard instrumental variable assumptions. It uses a method that is more robust to potential violations of the instrumental variables test result was noteworthy. This method examines the difference between the estimates from linear regression (observational) and instrumental variable analysis.

For binary outcomes, the instrumental variable analysis was done in two stages. Firstly, we assessed the association between the height or BMI genetic risk score and height or BMI, respectively. We saved the predicted values and residuals from this regression model. Secondly, we used the predicted values from stage 1 as the independent variable (reflecting an unconfounded estimate of variation in BMI or height) and degree status or job class as the dependent variable in a logistic or ordinal logistic regression model. We used robust standard errors to correct for uncertainty in the estimate. We examined the F statistics from first stage regressions to evaluate the strength of the instruments; weak instruments can bias results towards the (confounded) multivariable regression association or towards the null in a two stage design.

Results
Table 1 summarises the demographics of the 119,669 UK Biobank participants with valid genetic data and BMI and height measures. The height and BMI genetic risk scores were robustly associated with height and BMI (table 1). The association between the socioeconomic status measures and health outcomes and the associations between known height variants and height and known BMI variants and BMI in the UK Biobank are summarised in supplementary tables A-C.

Relation of genetically determined taller stature to higher socioeconomic status measures in UK Biobank

**Education: duration in full time education**
Using 82,543 participants, we found that taller stature was strongly correlated with participants spending longer in full time education (table 2). This association was similar in men and women. A 1 SD (6.3 cm) greater height was associated with a 0.11 (95% confidence interval 0.10 to 0.12) SD older age (approximately 0.2 years) at which full time education was completed. Genetic analyses provided evidence that this association was partly causal—a genetically determined 1 SD (6.3 cm) higher height was associated with a 0.03 (0.01 to 0.05) SD older age (approximately 0.06 (0.02 to 0.10) years) at which full time education was completed (table 2, supplementary figure A-I).

**Education: degree level (or equivalent) or not**
Using 118,565 participants, we found that taller stature was strongly correlated with participants’ chances of having obtained a degree. A 1 SD (6.3 cm) greater height was associated with 1.25 (1.28 to 1.27) increased odds of reporting degree level education. This association was similar in men and women (P for comparison>0.05). Genetic analyses provided no consistent evidence for a causal role of height in obtaining degree level education (table 2, supplementary figure A-II).

**Job class**
Using 76,404 participants, we found that taller stature was strongly correlated with job class. A 1 SD (6.3 cm) greater height was associated with increased odds of
working in skilled job roles (odds ratio 1.29, 1.27 to 1.32). Genetic analyses provided evidence that this association was partly causal—a 1 SD (6.3 cm) genetically determined higher height was associated with increased odds of working in more professional roles (odds ratio 1.12, 1.07 to 1.18) (table 2, supplementary figure A-III). This association was consistent when we analysed the data as 11 ordered job classes (supplementary table D). We found no genetic evidence that the effect was stronger in men or women.

**Annual household income**

Using 103 327 participants, we found that taller stature was strongly correlated with higher household income. The correlation was approximately 50% stronger in men (table 2). A 1 SD (6.3 cm) greater height was associated with a 0.13 (0.12 to 0.14) SD increase in income (table 2). This difference is approximately equivalent to a £2940 (£2730 to £3185) higher annual household income. Genetic analyses provided evidence that this association was partly causal—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, equivalent to £1130 (£680 to £1580) (table 2 and supplementary figure AIV). The genetic analyses showed that the effect was approximately twice as strong in men as in women (P for comparison=5×10⁻⁶); a 1 SD greater height in men caused a £1580 (£1130 to £2260) increase in household income (supplementary figure A-IV). This association was consistent when we analysed the data as five ordered income classes (supplementary table D).

**Townsend deprivation index**

Using 119 519 participants, we found that taller stature was strongly correlated with lower levels of social deprivation, as measured by the Townsend deprivation index. This association was stronger in men than women. A 1 SD (6.3 cm) greater height was associated with a 0.08 (0.07 to 0.09) SD lower deprivation, which is equivalent to a 0.21 (0.18 to 0.24) unit reduction in Townsend deprivation index (table 2). Genetic analyses provided evidence that this association was partly causal in men but not when all participants or women were considered. In all participants, genetically determined height was not associated with deprivation (table 2, supplementary figure A-V). In men, a genetically determined 1 SD (6.3 cm) greater height was associated with a 0.02 (0.00 to 0.05) SD reduction in deprivation (supplementary figure AV). This difference is equivalent to a 0.05 (0.00 to 0.13) unit lower Townsend index.

**Relation of genetically determined higher BMI to reduced income and deprivation measures of socioeconomic status in UK Biobank**

**Education: duration in full time education**

Using 82 543 participants, we found that higher BMI was strongly correlated with participants finishing full time education at a younger age. The association was similar in men and women (P for comparison>0.05)
Table 3 | Associations between higher BMI and five measures of socioeconomic, using linear or logistic regression and instrumental variable analysis

<table>
<thead>
<tr>
<th>Socioeconomic status measures and subcategories</th>
<th>Observational* Change in socioeconomic status (95%CI) per SD higher BMI</th>
<th>P value</th>
<th>Genetic† Change in socioeconomic status (95%CI) per SD higher BMI</th>
<th>P value</th>
<th>Genetic: Egger‡ Change in socioeconomic status (95%CI) per SD higher BMI</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age completed full time education:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>82543</td>
<td>−0.08 (-0.08 to −0.07)</td>
<td>&lt;1×10⁻¹⁵</td>
<td>−0.01 (−0.07 to 0.04)</td>
<td>0.63</td>
<td>−</td>
</tr>
<tr>
<td>Men only</td>
<td>38542</td>
<td>−0.07 (−0.09 to −0.06)</td>
<td>&lt;1×10⁻¹⁵</td>
<td>0.00 (−0.09 to 0.09)</td>
<td>0.98</td>
<td>−</td>
</tr>
<tr>
<td>Women only</td>
<td>44201</td>
<td>−0.08 (−0.09 to −0.07)</td>
<td>&lt;1×10⁻¹⁵</td>
<td>−0.02 (−0.09 to 0.05)</td>
<td>0.56</td>
<td>−</td>
</tr>
<tr>
<td>Degree level education:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>118565</td>
<td>0.04 (0.82 to 0.84)</td>
<td>&lt;1×10⁻¹⁵</td>
<td>0.94 (0.85 to 1.03)</td>
<td>0.18</td>
<td>−</td>
</tr>
<tr>
<td>Men only</td>
<td>56111</td>
<td>0.04 (0.82 to 0.84)</td>
<td>&lt;1×10⁻¹⁵</td>
<td>0.94 (0.81 to 1.09)</td>
<td>0.43</td>
<td>−</td>
</tr>
<tr>
<td>Women only</td>
<td>62454</td>
<td>0.04 (0.82 to 0.84)</td>
<td>&lt;1×10⁻¹⁵</td>
<td>0.93 (0.82 to 1.06)</td>
<td>0.28</td>
<td>−</td>
</tr>
<tr>
<td>Job class (skilled/unskilled):</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>76404</td>
<td>0.01 (0.89 to 0.92)</td>
<td>&lt;1×10⁻¹⁵</td>
<td>0.90 (0.79 to 1.02)</td>
<td>0.10</td>
<td>−</td>
</tr>
<tr>
<td>Men only</td>
<td>37608</td>
<td>0.01 (0.89 to 0.95)</td>
<td>&lt;1×10⁻¹⁵</td>
<td>0.88 (0.73 to 1.08)</td>
<td>0.22</td>
<td>−</td>
</tr>
<tr>
<td>Women only</td>
<td>38796</td>
<td>0.01 (0.89 to 0.91)</td>
<td>&lt;1×10⁻¹⁵</td>
<td>0.91 (0.76 to 1.06)</td>
<td>0.29</td>
<td>−</td>
</tr>
<tr>
<td>Annual household income:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>103327</td>
<td>−0.06 (−0.06 to −0.05)</td>
<td>&lt;1×10⁻¹⁵</td>
<td>−0.05 (−0.10 to −0.00)</td>
<td>0.041</td>
<td>−0.03 (−0.11 to 0.05)</td>
</tr>
<tr>
<td>Men only</td>
<td>50862</td>
<td>−0.01 (−0.02 to −0.00)</td>
<td>&lt;1×10⁻¹⁵</td>
<td>0.06 (−0.02 to 0.14)</td>
<td>0.15</td>
<td>−</td>
</tr>
<tr>
<td>Women only</td>
<td>52465</td>
<td>−0.09 (−0.10 to −0.08)</td>
<td>&lt;1×10⁻¹⁵</td>
<td>−0.16 (−0.20 to −0.08)</td>
<td>1×10⁻⁵</td>
<td>−0.17 (−0.25 to −0.05)</td>
</tr>
<tr>
<td>Townsend deprivation index:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>119519</td>
<td>0.08 (0.07 to 0.08)</td>
<td>&lt;1×10⁻¹⁵</td>
<td>0.05 (0.01 to 0.10)</td>
<td>0.24</td>
<td>−0.00 (−0.08 to 0.08)</td>
</tr>
<tr>
<td>Men only</td>
<td>56582</td>
<td>0.05 (0.04 to 0.05)</td>
<td>&lt;1×10⁻¹⁵</td>
<td>−0.01 (−0.08 to 0.06)</td>
<td>0.78</td>
<td>−</td>
</tr>
<tr>
<td>Women only</td>
<td>62937</td>
<td>0.10 (0.09 to 0.11)</td>
<td>&lt;1×10⁻¹⁵</td>
<td>0.10 (0.04 to 0.16)</td>
<td>0.001</td>
<td>0.10 (−0.01 to 0.21)</td>
</tr>
</tbody>
</table>

BMI=body mass index; OR=odds ratio.

*Age, assessment centre, and sex adjusted associations.

†Uses instrumental variable analysis, via ivreg2 command in Stata for continuous variables and two step approach for binary outcomes, using BMI genetic risk score. F statistic for all associations between higher BMI and five measures of socioeconomic status represents odds of higher socioeconomic status per SD higher BMI.

‡Alternative genetic approach detailed in Bowden et al 2015, 24 used as sensitivity analysis when instrumental variable was P status measure.

Annual household income

Using 103 327 participants, we found that higher BMI was associated with a lower annual household income, but this effect was very strongly driven by the association in women. A 1 SD higher BMI was associated with a 0.09 (0.08 to 0.10) SD lower household income for women. This effect equates to £1890 (£1680 to £2100) less income per annum for women. In men, a 1 SD higher BMI approximated to a £210 (£84 to £420) lower annual household income. Genetic analyses were consistent with these observations being causal in women but not in men (P for comparison with men=9×10⁻³)—a genetically determined 1 SD higher BMI was associated with an annual household income of 0.14 (0.08 to 0.20) SD less in women. This effect is equivalent to £2940 (£1680 to £4200) less for women (table 3, supplementary figure B-IV). This association was consistent when we analysed the data as five ordered income classes (supplementary table D). The association between higher BMI and lower income was consistent in women who worked, with or without a husband/partner at home, and women who did not work with a husband/partner at home (supplementary table E). It was also consistent when we considered only women without health conditions (supplementary table E).

Townsend deprivation index

Higher BMI was associated with higher levels of deprivation as assessed by the Townsend deprivation index. A 1 SD higher BMI was associated with a 0.08 (0.07 to
0.08) SD higher deprivation value, which is equivalent to a 0.21 (0.19 to 0.21) unit increase in Townsend index (table 3). This association was twice as strong in women. We found limited genetic evidence of a causal relation between BMI and deprivation in men, but some evidence in women. A 1 SD genetically higher BMI was associated with a 0.10 (0.04 to 0.16) SD higher level of deprivation in women (table 3, Supplementary figure B-V).

Sensitivity analyses
The Egger method provided consistent results for causal relations between height and duration in full time education, job class, income, and Townsend deprivation index in men (table 2, supplementary table F). The Egger method also provided evidence of consistent associations between higher BMI and income in women (table 3, supplementary table G). Use of genome-wide methods to account for genetic and socioeconomic status correlations between close and distant relatives did not alter our findings (supplementary table H).

Discussion
Using genetic variants as unconfounded proxies for height and BMI, our study provides evidence that shorter stature and higher BMI lead to lower measures of several aspects of socioeconomic status. It is important to note that our data are consistent with the height and BMI to socioeconomic status associations being only partly causal—we have not excluded a causal effect in the other direction. The study adds causal evidence to a large number of observational studies. This work may have important implications for public health, as low socioeconomic status increases mortality and morbidity.23 The association between socioeconomic status measures and health was strong in the UK Biobank data, where, for example, people possessing a university degree had a 38% lower odds of coronary artery disease compared with those without degree level education. Our study also showed sex differences in the causal relations between height or BMI and socioeconomic status that are consistent with observational data. Height effects were stronger in men, but the BMI effects tended to be stronger in women.

Evidence for taller stature leading to higher socioeconomic status
The causal effect, as estimated using genetics, of taller stature on higher socioeconomic status was present in four of the five measures of socioeconomic status. For income, where the statistical evidence was strongest, the estimated causal effects were approximately two to three times stronger in men than in women. The causal evidence for greater height leading to higher levels of socioeconomic status is consistent with observational studies, in which taller stature was associated with higher job class, earnings, and educational attainment.35 One US based study showed a reduction in earnings of $789 per annum per inch of height. This equates to £1250 per SD (6.3 cm) of height in our study, which is very similar to our genetic estimate of £1130.35 Despite the strong evidence that taller stature directly influences measures of socioeconomic status, the genetic estimates were consistently smaller than the observational estimates. These differences indicate that the observed association between taller stature and higher socioeconomic status is a mixture of direct causal effects and other factors that could include a causal effect in the opposite direction.

A range of factors could link taller stature to higher socioeconomic status, although this study did not consider which of these factors were involved. Some of the possibilities include complex interactions between self esteem, stigma, positive discrimination,13 and increased intelligence.27 28 Evidence shows that self esteem, leadership perception, and height discrimination tend to be greater in men than in women, which fits with our findings.29 31

Evidence for higher BMI leading to lower socioeconomic status
Higher BMI, as estimated using genetics, was causally associated with having a lower annual household income and higher levels of deprivation. These associations were stronger in women, with no consistent evidence of a causal relation between higher BMI and lower socioeconomic status measures in men. These findings were consistent with previous literature, in which most associations of BMI with socioeconomic status were observed in women only.7 32 We found no evidence that the associations between higher BMI and educational outcomes were causal, a result consistent with a review of the effect of BMI on social outcomes.7 32 Our findings add to evidence from observational studies, in which BMI is associated with lower levels of employment, less skilled work, and lower income.32 33 A range of factors could link higher BMI to lower income and higher deprivation in women, although this study did not consider which of these factors were involved. One of the possibilities is discrimination in the workplace, with overweight job applicants and employees being evaluated more negatively.32 The disparity between the sexes may be partially explained by discrimination, which may occur at lower weight levels for women than for men.34 35 Additionally, cultivation theory in social science indicates that very thin women are idealised and more socially valued, compared with their normal weight and overweight peers.31 In contrast, a very different set of social standards exists regarding men’s weight, so discrimination based on body size could well be different in men and women.33 Two of the strongest measures in women were household income and Townsend deprivation, which are not just specific to the individual but also indicative of partner’s income. However, additional analyses showed that genetically determined higher BMI was associated with lower income both in non-working women with partners and in working women without a partner, suggesting that the associations were not just driven by partner’s income.

Limitations of study
Although our results are consistent with a direct causal effect of shorter stature and higher BMI on lower
socioeconomic status, some qualifications should be considered. Firstly, the UK Biobank participants were born between 1938 and 1971, and the causal associations may not remain in today’s society or be generalisable to societies outside of the United Kingdom. The causal associations may have been influenced by parental genotype-socioeconomic status associations. For example, the causal pathway could reflect parental genetic predisposition to higher BMI resulting in families moving to a more obese and lower socioeconomic status neighbourhood, which in turn could lower children’s socioeconomic status. Because parents’ and children’s genotypes are correlated, this pathway could lead to a genetic association between UK Biobank participants’ socioeconomic status and BMI that reflects parental factors during the 20th century. However, such a pathway would be unlikely to result in genetic associations between BMI and socioeconomic status that were stronger in women than in men.

Secondly, higher BMI leads to poorer health; this could affect productivity, which in turn could affect socioeconomic status. However, we saw evidence of genetic associations between higher BMI and lower socioeconomic status in women reporting no adverse health outcomes as well as in those reporting health problems (supplementary table E). We also need to take care in interpreting negative results; although the large sample size of the UK Biobank provided greater than 95% power for investigating the causal relations between height and socioeconomic status, power was limited for some of the causal associations for BMI. The SNPs selected for height and BMI may have effects on socioeconomic status not mediated by their effects on height or BMI (pleiotropy), which were not measured but could potentially affect socioeconomic status. However, to minimise this possibility, we selected SNPs carefully and used the Egger method, which can detect and adjust for pleiotropy bias in many scenarios (hence the broader confidence intervals observed).

The educational, job status, and income data used in this study were self-reported, which may result in measurement bias. However, Townsend deprivation index was derived by the UK Biobank and we observed consistent trends across the different socioeconomic status constructs, suggesting limited bias due to self-report. Socioeconomic status is a very complex multidimensional construct. We looked at a range of individual components and observed similar trends for each, but the selected variables may not cover the entirety of social status. This study used a homogenous population, so the results may not be generalisable to other ethnic groups. We note that the genetic variants associated with BMI together explain only 1.5% of the variation in BMI, but collectively the variants provide a robust test, as reflected by the strong F statistic of the genetic risk score. The 69 genetic variants associated with BMI provided a stronger “instrument” than those used in previous mendelian randomisation studies that inferred a causal effect of higher BMI leading to lower vitamin D concentrations and higher risk of heart failure and markers of poor metabolic health.36 37

Finally, height, BMI, and socioeconomic status are subtly stratified across the United Kingdom, with people living and working in the north of the country having lower socioeconomic status, higher BMI, and shorter stature, on average, than those in the south. If genetic variants are also subtly different between north and south, this could have confounded our results. However, several factors mean that this population stratification should not have caused false positive results. Firstly, we would not have expected to have seen differences between men and women (because gene allele frequencies do not differ between the sexes). Secondly, we used both within UK genetic ancestry principal components and a second method that corrects for all levels of relatedness, and our results did not change.

Conclusion
In summary, using up to 119,000 participants from the UK Biobank, we provide evidence that high BMI and short stature, as estimated by genetics, are causally related to lower socioeconomic status. Further work is needed to understand the factors that lead to and from anthropometric traits and socioeconomic status.

This research has been conducted using the UK Biobank resource.

Contributors: JT, TMF, and MNW designed the study and wrote the manuscript. CMA and JNH edited the manuscript and helped to interpret the data. SEJ, JT, RB, KSR, ARW, MAT, HY, RMF, AM, and MNW were involved in data processing, statistical analyses, and interpretation. TMF is the guarantor.

Funding: JT is funded by a Diabetes Research and Wellness Foundation fellowship. SEJ is funded by the Medical Research Council (grant: MR/M005070/1). MAT, MNW, and AM are supported by a Wellcome Trust institutional strategic support award (WT097835MF). ARW and TMF are supported by a European Research Council grant (7245 59371-GLUCOSEGENES-FP-IDEAS-ERC). RMF is a Sir Henry Dale fellow (Wellcome Trust and Royal Society grant 104150/2/14/2). RB is funded by a Wellcome Trust and Royal Society grant (104150/2/14/2). HY is funded by a European Research Council award (333195). JNH is supported by the National Institutes of Health (R01 DK075787). CMA is funded by the Endocrine Society’s Endocrine Scholars Award (supported by Lilly USA, LLC). The funders had no influence on study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organisation for the submitted work other than detailed above; MNW has received speakers fees from Ipsen and Merck, and TMF has consulted for Boehringer Ingelheim; no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: UK Biobank has received ethics approval from the National Health Service National Research Ethics Service (ref 11/ NW/0382).

Transparency statement: The lead author (the manuscript’s guarantor) affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Data sharing: The data reported in this paper are available via application directly to the UK Biobank.

This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 3.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: http://creativecommons.org/licenses/by/3.0/.


