I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17699

Returns to Testosterone Across Men's Earnings Distribution in the UK

Peter Eibich
Ricky Kanabar
Alexander Plum

FEBRUARY 2025

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17699

Returns to Testosterone Across Men's Earnings Distribution in the UK

Peter Eibich
Université Paris Dauphine-PSL and IZA

Ricky Kanabar
University of Bath

Alexander Plum
Auckland University of Technology

FEBRUARY 2025

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Returns to Testosterone Across Men's Earnings Distribution in the UK*

We study how population variation in testosterone levels impacts male labour market earnings using data from the UK Household Longitudinal Study between 2011 and 2013. We exploit genetic variation between individuals as instrumental variables following a Mendelian Randomization approach to address the endogeneity of testosterone levels. Our findings show that higher testosterone levels have a strong positive impact on earnings. Importantly, these findings are limited to men belonging to the lower quartile of the testosterone distribution and working in higher-paid jobs. We show that differences within rather than between occupations drive these findings, whereas we find limited support for selection into occupation or mechanisms involving individual characteristics, including personality traits and education.

JEL Classification: J31, C26, I10
Keywords: earnings, IV, testosterone, Mendelian Randomisation

Corresponding author:
Ricky Kanabar
Department of Social Policy and Social Sciences
University of Bath
Claverton Down
Bath BA2 7AY
Great Britain
E-mail: rk735@bath.ac.uk

* The data used in this paper are restricted access; individuals interested in accessing the data can contact the corresponding author for details. We would like to thank Anna Dearman for assistance with generating the polygenic score, and Matt Dickson, Amanda Hughes and Thomas Schober for their helpful comments and suggestions on an earlier draft of the paper.

1 Introduction

Testosterone has been linked to social and economic outcomes in studies dating back to the 1970s (Dabbs, 1992; Kreuz and Rose, 1972; Mazur and Lamb, 1980). Testosterone affects physical development (e.g., the development of male characteristics in utero or during puberty, muscle mass), but also influences behavioural responses (Phoenix et al., 1959). For example, in humans, testosterone has been linked to status-enhancing behaviour (Dreher et al., 2016), such as aggressiveness (Carré and McCormick, 2008), risk-taking (Apicella et al., 2008; Coates and Herbert, 2008) but also prosocial behaviour (Dreher et al., 2016; Schaal et al., 1996).

This latter association between testosterone and behaviour, in particular, inspired a strand of the literature examining the role of testosterone for labour market outcomes (Dabbs, 1992; Dabbs Jr. et al., 1990). For example, studies report that higher testosterone is associated with higher rates of self-employment (Greene et al., 2014; Nicolaou et al., 2017), careers in "risky" occupations such as finance (Nye and Orel, 2015; Sapienza et al., 2009), and higher earnings (Gielen et al., 2016; Hughes and Kumari, 2019; Nye et al., 2017).

The causality of these findings is unclear. Testosterone fluctuates substantially - over the course of a day, across socioeconomic transitions such as marriage or divorce (Mazur and Michalek, 1998), and over the life course (Kanabar et al., 2022). The causes of these fluctuations are not fully understood, and studies disagree on the extent to which testosterone activates or reacts to particular behaviour. For example, the challenge hypothesis posits that testosterone levels increase when males[1] have to compete with others, and this rise in testosterone levels activates behaviour that allows individuals to succeed (Archer, 2006). At the same time, there is substantial evidence that the outcome of a competition affects testosterone levels (Geniole et al., 2017), even when the outcome is entirely determined by chance (McCaul et al., 1992). Consequently, whether testosterone contributes to better labour market outcomes or economic success increases testosterone levels remains an open question.

[1] The hypothesis was originally proposed to explain the behaviour of male birds, but has subsequently been applied to humans and other animals (Wingfield, 2017).

This study examines the effect of testosterone on wages among British men. We use genetically enhanced data from the "Understanding Society" (UKHLS) study, which allows us to construct a polygenic score measuring men's genetic predisposition towards higher testosterone levels. We use this polygenic score as an instrumental variable to address the endogeneity of testosterone. Our results show that a higher testosterone level increases men's earnings.

To identify the causal effect of testosterone, previous economics studies have relied on instrumental variables, for example, biomarkers such as insulin (Greene et al., 2014; Nicolaou et al., 2017). The 2D:4D ratio[2] has been used as a proxy for prenatal testosterone exposure (Nye et al., 2017; Nye and Orel, 2015), but it is well-established that it does not predict hormone levels in adults (Hönekopp et al., 2007; Kowal et al., 2020; Zhang et al., 2020). Gielen et al. (2016) also consider the effects of prenatal testosterone exposure using twinning as a natural experiment.

The present study considers the effects of circulating testosterone measured in adult men. The organizational-activational hypothesis (Arnold, 2009; Phoenix et al., 1959) posits that testosterone exposure during critical periods (e.g., in utero) has permanent effects on the development of tissue, including the brain ("organizational effects"). In adulthood, testosterone exposure then activates certain behavioural responses (Arnold, 2009). This implies that the effects of prenatal exposure may differ from the effects of exposure to circulating testosterone in adulthood, as the former primarily captures organizational effects and the latter focuses on activational effects.[3]

Two recent studies also exploit genetic variants as instruments for circulating testosterone. Hughes and Kumari (2019) draw on the same UKHLS data used in our analysis and report a marginally significant increase in gross earnings with testosterone. In contrast, Harrison et al. (2021) find no significant effects on a substantially larger sample from the UK Biobank.

We move beyond these studies by considering nonlinear effects of testosterone exposure. We use our instrument to estimate control function regressions (Wooldridge, 2015), which allow us to

[2] The 2D:4D ratio measures the relative length of the 2nd and the 4th finger ("digit").
[3] It also implies that, ideally, prenatal exposure and exposure to circulating hormone levels in adulthood should be studied together. Unfortunately, our data does not allow us to measure prenatal exposure.

flexibly model nonlinear trends in the endogenous variable. Our results show that the positive effects of testosterone are concentrated at the lower end of the testosterone distribution as well as at the higher end of the earnings distribution. In other words, testosterone increases the likelihood of men with low testosterone levels to belong to the group of top earners, but we observe little to no effects among men with medium or high levels of testosterone, as well as men earning low wages.

We further contribute to the literature by examining potential mechanisms that might explain the testosterone wage premium. First, we show that selection into occupations and job characteristics play at best, a minor role. We find little evidence that testosterone affects occupational choice or job characteristics such as working hours or performance pay. Moreover, accounting for occupation-fixed effects explains less than 20% of the total effect of testosterone on wages. Next, we consider a wide range of individual characteristics that have been linked to testosterone in previous studies. We find no effects of testosterone on educational attainment, cognitive functioning or the Big 5 personality traits. Higher testosterone levels significantly increase risk tolerance; however, accounting for risk tolerance does not seem to explain the positive effect of testosterone on wages.

There are several potential mechanisms that we are not able to examine empirically due to a lack of data. For example, it is possible that testosterone levels are linked to occupational tasks or skills. It is also possible that selection into occupation occurs within the broader occupational groups considered here. Testosterone may also influence men's behaviour in specific situations, such as salary negotiations, which we cannot observe in social surveys.

Regardless of the precise mechanisms, our results suggest that biological factors (such as hormones) are essential in determining labour market outcomes. A better understanding of the mechanisms behind these effects may allow us to determine to which extent such premiums are warranted or whether they reflect discrimination (e.g., on physical features; Hammermesh and Biddle, 1994) that might result in labour market inefficiencies.

The remainder of this paper is structured as follows: Section 2 describes the data we used and how we derived the polygenic scores; Section 3 presents the methodological approach, followed by a description of the results in Section 4. Section 5 concludes.

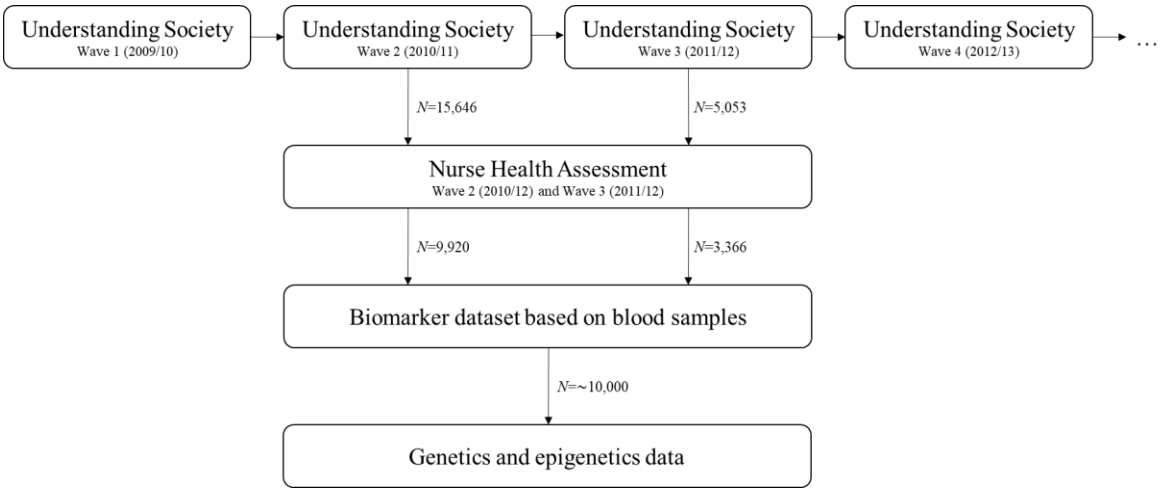
2 Data

2.1 The Understanding Society Survey

Our study uses data from the longitudinal survey Understanding Society (or UK Household Longitudinal Study, UKHLS). The UKHLS is the successor of the British Household Panel Survey, which began in 2009 by interviewing about 40,000 households and around 100,000 individuals (ISER, 2023*a*). The annual survey provides an extensive range of economic and socio-demographic information for individuals and households.

At waves 2 and 3, about 20,000 participants (excluding those residing in Northern Ireland) were invited for the so-called Nurse Health Assessment (Benzeval et al., 2014); see also Figure 1. The nurse visit took place approximately five months after the main wave interview, collecting information on the individual's medical history as well as physical measures, such as height, weight, lung function, blood pressure, and grip strength. Blood samples were collected from a subset of about 13,000 consenting individuals during the nurse visit. Based on the blood samples, several biomarkers have been extracted, including measures of growth hormones (e.g., testosterone, DHEA's, IGF-1). Finally, a genome-wide scan of approximately 10,000 eligible and consenting individuals with a blood sample was conducted on DNA samples.

Figure 1: The dataset structure



Source: Authors own representation and Understanding Society (see https://www.understandingsociety.ac.uk/topic-page/biomarkers-genetics-and-epigenetics/)

2.2 Working sample

UKHLS contains information on the respondent's current labour force status, earnings, hours worked (including normal and overtime hours), and type of occupation. We restrict our working sample to men aged 25-64 during the nurse visit.[4] We trim the sample to employees working full-time, and whose monthly earnings are at least £900 (2015 prices), the equivalent of earning minimum wage at the time of the survey interview. Using the same data and a similar econometric approach, we previously documented that testosterone is linked to labour force participation at the extensive margin (Eibich et al., 2022). In this study, we therefore focus on earnings at the intensive margin to complement our earlier findings.

We drop self-employed individuals since this group of individuals is likely to differ from employees in terms of their unobservable characteristics, such as personality type, as well as their labour supply behaviour. Our sample is therefore comprised of individuals for whom testosterone does not influence their decision to take up self-employment. Our primary outcome of interest is an individual's total monthly gross labour market earnings, which have been log-transformed and

[4] We do not study the relationship between earnings and testosterone levels for women as for most women in the UKHLS sample testosterone levels are below the detectable threshold.

deflated using the Consumer Price Index inclusive of Housing costs (CPIH, figures correspond to 2015 prices; see ONS, 2024). We focus on total log earnings as opposed to income from all sources which may include inter-alia unemployment benefit given our aim is to study the effects of testosterone and its relationship with outcomes produced in a competitive labour market setting.[5] After imposing these sample restrictions and accounting for data cleaning, our base sample consists of 1,621 individuals with a complete information set.

2.3 Nurse Health Assessment and Biomarker Dataset

We utilise information from the nurse visit stored in the Nurse Health Assessment dataset, including the time the interview took place during the day. Serum testosterone, the specific biomarker of interest for this study, was measured using an electrochemiluminescent immunoassay on the Roche Modular E170 analyser. The steroid is calculated in nanomoles per litre (nmol/l). Testosterone levels vary across men and are considered normal within a range between 9-25 nmol/l. Testosterone varies by time of day, such that values in the morning are higher than those found in the afternoon or evening (see Table 1 of Eibich et al., 2022). For our analysis, we retain individuals whose nurse interview occurred between 09:00 and 20:00.[6]

Individual's testosterone level also declines with age (Kanabar et al., 2022). To adjust for these diurnal patterns and age trend, we create 5-year age bands (25-29, 30-34, etc.) and estimate a separate regression for each age group in which we regress testosterone on the interview time (centred around 10:00). For each individual we calculate the deviation from the predicted interview-time corrected mean testosterone level of the corresponding age group. Finally, we standardise our adjusted measure of testosterone by dividing our estimate by the age-band-specific standard deviation.

[5] It is likely the mechanisms which link testosterone and total income are distinct to those for earnings.
[6] Around 2% of the survey participants had their interview outside this time window. Roughly half of them had their interview between 01:00 and 02:00. Specific care must be undertaken when including these persons as their testosterone level is measured before their sleep, resulting in lower testosterone levels compared to those interviewed a few hours later.

2.4 Genetic data

We draw on genetic data collected during the Health and Biomarker Survey to address the endogeneity between earnings and testosterone. Access to the data was granted by the UKHLS Genetics Committee. Our approach is to investigate whether genetic variants that partly explain the variance of serum testosterone affect labour market earnings. This method, which uses genetic markers as so-called Instrumental Variables (IV) is known as Mendelian Randomisation (MR) (Burgess and Thompson, 2015) and is outlined formally in Section 4.

Genetic variation arises due to single nucleotide polymorphisms (SNPs), where each SNP represents a difference in a single DNA building block (i.e., a nucleotide) and is randomly assigned at conception. We use the genetic data to construct so-called polygenic scores (PGS).[7] A PGS provides an estimate of an individual's genetic predisposition to a particular trait (i.e., phenotype) and is derived based on their genotype profile, combined with effect sizes calculated from a Genome Wide Association Study (GWAS). To construct these scores, which typically sum multiple SNPs, we follow the classic approach, commonly referred to as Clumping and Threshold ('C+T'), which allows for effect sizes (multiplied by individual's respective SNP) to be summed together and combined with selection based on p-value (Choi et al., 2020). For analysis purposes, all PGS are standardised to facilitate interpretation.

The first PGS is based on the GWAS conducted by (Ohlsson et al., 2011), which identified three genetic variants to use as instruments. These are rs12150660 and rs6258 in the SHGB gene on chromosome 17 and rs5934505 near FAM9B on the X chromosome.[8] In the UKHLS data, rs12150660 was imputed, whereas rs6258 and rs5934505 were genotyped (Hughes and Kumari, 2019). Such imputation is typical given the resources required to sequence SNPs, therefore in studies using a MR approach which involve multiple SNPs to generate an adequately powered PGS unmeasured SNPs are imputed based on a reference dataset taking into account possible

[7] Individual's genetic data were genotyped using the Illumina HumanCore Exome, and imputation was carried out in Minimac 5-12-29 to the European component of 1000 genomes (Hughes and Kumari, 2019). Samples are checked to ensure genetic data is consistent with key information provided, such as gender and ethnicity. Quality control checks removed SNPs with a minor allele frequency of <1%, call rate threshold <98%, Hardy-Weinberg Equilibrium $p<10^{-4}$, or cluster separation score <0.4 (Hughes and Kumari, 2019).

[8] The effect allele in the Ohlsson et al. (2011) study refers to the minor allele (A1). The dosages provided in UKHLS refer to those recorded in A2. Dosages are adjusted (flipped) accordingly.

linkage to measured SNPs (see Benzeval et al., 2014). Imputation is not based on testosterone or any other factor related to our outcome of interest. Ohlsson et al. (2011) and Hughes and Kumari (2019) show that the single combined polygenic score explains approximately 3% of the variance in circulating serum testosterone. Working with the UKHLS Genetics team, we created an updated version of the polygenic score using the same genetic variants that were used by Hughes and Kumari (2019).

The second polygenic score is based on the GWAS conducted by Harrison et al. (2021) using the UK Biobank. In this study, for men a much larger number of SNPs (70) were identified that held genome-wide significance with respect to total testosterone. These SNPs jointly explained 6.83% of the variation in total testosterone. At the time of writing, UKHLS only contained a measure of total (as opposed to bioavailable) testosterone, which we used for analysis purposes.[9] Whilst not all SNPs identified are available in the UKHLS data, 51 (approximately 73%) were. Additionally, we were able to proxy a further 9 SNPs (subject to a minimum correlation R^2 of 0.8), allowing us to create a polygenic score based on this much larger GWAS study.[10,11,12] However, we note that the most powerful single SNP identified by Ohlsson et al. (2011), rs12150660, which explains 1.7% of circulating total testosterone alone, is not available in the UK Biobank and thus not used in the creation of the PGS that follows the Harrison et al. (2021) study. This, combined with the fact that not all SNPs identified in Harrison et al. (2021) are available in UKHLS, means that the association between the PGS score and testosterone (i.e., the first-stage regression) in our study is smaller than the association (and hence explained variation in testosterone) reported in the original study and that based on the PGS developed by Hughes and Kumari (2019), as shown in Table 2.[13]

[9] Whilst total and bioavailable testosterone (loosely bound to albumin or tightly bound to SHGB) are highly correlated, the latter is available for biological processes and hence is perceived as providing a better construct for determining causal processes (Harrison et al., 2021).

[10] We account for Linkage Disequilibrium (LD) in the context of proxy SNPs by checking which allele in the proxy SNP is in LD with the effect allele in the GWAS.

[11] One must determine the dosage of the effect allele for each SNP within the PGS. In some cases, this is a partial number due to imputation (in the case of deriving the SNP based on Harrison et al. (2021)), and we retain imputed values to account for this uncertainty.

[12] The list of SNPs used to derive the PGS based on the Harrison et al. (2021) study is available on request.

[13] It is not possible to combine PGS or include both instruments for regression purposes due to the risk of linkage between genetic variants in both PGSs, leading to violations of the exclusion restriction.

3 Descriptive Statistics

Table 1 presents descriptive statistics of key socioeconomic and demographic characteristics based on our estimation sample. The average age among males at the time of the respective nurse and mainstage interview is 45, and total gross monthly labour earnings are close to £3,000 (2015 prices). Approximately one third of the sample reports working in a managerial or professional occupation and two fifths hold a degree. Roughly 65% are legally married, and one in five males reports their main residence as being in London or the South East.

Table 1: Descriptive statistics at the time of nurse and mainstage interview

		Testosterone level (binary)		
	Full Sample	Medium/high	Low	*t*-test
Age	44.64	44.76	44.19	0.914
	(10.21)	(10.27)	(9.98)	(0.361)
Total earnings (in £)‡	2963.43	3021.48	2737.20	3.018***
	(1532.53)	(1576.31)	(1326.51)	(0.003)
Occupation: ISCO 1 or 2	0.367	0.373	0.344	0.958
(Managers/Professionals)	(0.482)	(0.484)	(0.476)	(0.338)
Degree (incl. other	0.408	0.411	0.396	0.498
higher)	(0.492)	(0.492)	(0.490)	(0.619)
Married	0.646	0.637	0.680	-1.444
	(0.478)	0.481	0.467	(0.149)
London / South East	0.200	0.203	0.187	0.640
	(0.400)	(0.402)	(0.391)	0.522
Personality traits				
Prepared to take risks†	5.826	5.874	5.636	1.281
	(2.331)	(2.350)	(2.251)	(0.200)
Cognitive numeric ability	4.136	4.155	4.062	1.623
(correct answers) †	(0.918)	(0.917)	(0.918)	(0.105)
Testosterone (nmol/l)	14.90	16.55	8.46	32.235***
	(5.22)	(4.43)	(2.15)	(0.000)
N	**1,621**	**1,290**	**331**	

Source: UKHLS, own calculations. *Notes*: † Refers to a subset of individuals. ‡ deflated using the Consumer Price Index inclusive of Housing costs (see ONS, 2024). Numbers in () are std dev and for the *t*-test the respective *p*-value. Significance of the *t*-test: *** p<0.01, ** p<0.05, * p<0.1

In Columns 3 and 4 of Table 1 we split the sample and define individuals as having low (medium/high) testosterone if their adjusted testosterone level belongs to the 1st or 2nd (3rd and

above) decile of the respective distribution.[14] When comparing the samples, we observe a significant difference in total gross monthly earnings which equates to around a £300 per month, approximately 10% of average earnings across the full sample. In the bottom part of Table 1, we include two markers of personality traits: risk-taking behaviour and cognitive numeric skills. In this case we do not observe a difference in average levels of either trait at conventional levels of significance.

4 Methods

4.1 Mendelian Randomization

Previous studies document that testosterone is associated with many characteristics that are in turn linked to income, e.g., occupation (Dabbs, 1992; Dabbs Jr. et al., 1990), risk-taking (Apicella et al., 2008; Coates and Herbert, 2008; Stenstrom et al., 2011), generosity (Ou et al., 2021), or cognition (Brosnan et al., 2011; Chance et al., 2000). While these characteristics may explain a hypothetical effect of testosterone on earnings, they also raise concerns about causality. Mazur and Michalek (1998) show that testosterone levels change following marriage or divorce, which more broadly suggests that social and economic factors influence testosterone as well. Thus, it is not clear to what extent some of the findings in this literature should be interpreted as causal effects of testosterone or whether they might reflect omitted variable bias or reverse causality.

Recent studies on testosterone and economic outcomes have addressed this endogeneity problem in different ways, e.g., with testosterone injections in a randomized experiment (Dreher et al., 2016), by using the 2D:4D ratio as a proxy for prenatal exposure to sex hormones (Neyse et al., 2021), or through Mendelian Randomization analysis (Eibich et al., 2022; Harrison et al., 2021; Hughes and Kumari, 2019).

Mendelian Randomization (MR) analysis uses genetic variants that are associated with a certain phenotype as instrumental variables (IV) for this phenotype. A person's genome is determined at

[14] Our main findings are robust to changing the threshold for defining whether an individual has low testosterone (we test a range of thresholds between the 15th and 30th percentile of the testosterone distribution; results available on request).

conception through random recombination of their parents' genes; therefore, the presence or absence of specific genetic variants in a person's genome is also random conditional on ancestry. Under the standard IV assumptions, estimating the causal effects of the relevant phenotype is possible. In the MR framework, these assumptions require that (1) the genetic variants are sufficiently strong predictors of the phenotype under study,[15] (2.a) the presence of these genetic variants can be considered as random conditional on parents' genes,[16] (2.b) the specific genetic variants are not associated with other phenotypes that might be related to the outcome under study, and (3) the direction of the association between genetic variants and phenotype is the same for all individuals in the sample. Under these three assumptions, IV models estimate a local average treatment effect. For MR analyses, this implies that we estimate the causal effect for individuals whose treatment status is determined by their genetic predisposition towards the phenotype, which is constant over a person's life course. We discuss the plausibility of the standard IV assumptions in more detail in Section 4.3.

4.2 Estimation

We estimate our IV regression models using control functions (Wooldridge, 2015). In the first stage of the model, we regress testosterone T_i on the polygenic score PGS_i while controlling for k different characteristics $x_{1i}, x_{2i}, \dots, x_{ki}$:

$$T_i = \beta_0 + \theta PGS_i + \sum_{j=1}^{k} x_{ji}\beta_j + e_i \quad (1)$$

Our baseline models control for the first 10 principal components of the genome of individual i and age group.[17] Controlling for the principal components is common practice in Mendelian

[15] Although the detection of relevant genetic variants in a genome-wide association study (GWAS) implies that there is a robust correlation between genetic variants and phenotype, it is possible that this correlation cannot consistently be replicated, e.g., because of differences in the composition of the populations of the MR sample and the GWAS sample, or because samples used for the GWAS tend to be much larger than samples used for other analyses, including MR analyses.

[16] This assumption is violated if testosterone affects parents' testosterone levels, which in turn influences parenting styles, and this is related to assortative mating in the parent generation. Such dynastic effects imply a relationship between the instrument and the outcome that does not work through the treatment which, if true, would violate the independence assumption.

[17] Although our measure of testosterone is already adjusted for age differences, we nevertheless control for age to account for the lifecycle profile of earnings in the second stage of the model.

Randomization analyses to account for population stratification (Davies et al., 2018). [18] Using the first stage regression in eq. (1), we predict residuals $\hat{e}_i$ to estimate the structural earnings equation:

$$\ln(inc_i) = \gamma_0 + \tau f(T_i) + \sum_{j=1}^{k} w_{ji}\gamma_j + \rho\hat{e}_i + u_i \quad (2)$$

In eq. (2), we regress log income on a flexible function of testosterone $f(T_i)$, covariates w_{ji} (identical to those included in the vector x_i in the first stage regression) and the predicted residual from the first stage $\hat{e}_i$. τ is a vector of one or more parameters (depending on the functional form of testosterone) that captures the causal effect of testosterone on earnings.

Intuitively, the first stage regression decomposes the observed variation in testosterone levels into a random part (explained by the PGS and the control variables), and an endogenous part (captured by the residual e_i), and we then control for the endogeneity of testosterone in the structural equation by including the predicted residual as a covariate. If we use the same functional form for testosterone in the first stage and the structural equation ($f(T_i) = T_i$, i.e., a linear trend) and we control for the same set of covariates in both regressions (i.e., $\{x_{1i}, x_{2i}, \dots, x_{ki}\} = \{w_{1i}, w_{2i}, \dots, w_{li}\}$), then the control function estimate will be identical to the 2SLS estimate.

An important advantage of control functions over 2SLS estimation is that the former allows us to flexibly consider different functional forms for testosterone in the structural equation (2). There is no reason to think that testosterone should have a linear effect on earnings. Assuming other functional forms in a 2SLS regression model would require constructing additional instruments for nonlinear terms (e.g., quadratic or cubic polynomials). While, e.g., quadratic terms of the PGS can serve as instruments for squared testosterone levels, these will likely be weak instruments because there is limited variation in the polygenic scores and consequently the PGS does not predict nonlinear terms in testosterone well (see Figure 2). Consequently, control function estimates of such models are likely more efficient (but potentially less robust) than 2SLS estimates (Wooldridge, 2015). Finally, the estimated parameter on the predicted residuals, $\hat{\rho}$, allows for a simple Hausman test of the endogeneity of testosterone – i.e., if $\hat{\rho}$ is statistically significant, we

[18] Intuitively, the principal components control for ancestry because certain genetic variants occur more frequently in some population groups.

can reject the hypothesis that testosterone is exogenous in the structural equation. We note that the usual standard errors of the estimates of the structural equation do not account for the estimation of $\hat{e}_i$ in the first stage regression. We, therefore, use bootstrapped standard errors based on 1,000 replications.

4.3 Plausibility of IV assumptions

The estimates of $\hat{\tau}$ in eq. (2) have a causal interpretation assuming the genetic variants we use in the Mendelian Randomisation approach satisfy the conditions for being valid instruments. Namely, the instruments should meet the following criteria: (1) relevance, (2a) independence, (2b) excludability and (3) monotonicity. Considering relevance, Figure 2 shows a positive association between an individual's adjusted testosterone level and their polygenic score. This is confirmed in the first stage regression results. Yet, we note that while our estimated first-stage effect is highly significant, our genetic variants only explain limited variation in testosterone and might suffer from weak instrument problems (Keane and Neal, 2023; Stock and Yogo, 2005). A weak first stage exacerbates the finite sample bias of 2SLS estimates and can invalidate statistical inference (Lee et al., 2022). It is common practice to screen for instrument relevance based on the sample F-statistic and various threshold values proposed in the literature (Montiel Olea and Pflueger, 2013; Stock and Yogo, 2005); however, Angrist and Kolesár (2024) show that in a just-identified IV model such pretesting is problematic, and screening on the estimated sign of the instrument is sufficient.

Keane and Neal (2023) show that inference based on 2SLS estimates can be invalid even if the sample F-statistic exceeds commonly used threshold values and recommend the use of alternative estimators (e.g., LIML) or tests (e.g., Anderson-Rubin) that are robust to weak instrument problems. Unfortunately, to the best of our knowledge, there is little guidance in the literature on addressing weak instrument problems in control function regressions. Burgess and Thompson (2012) show in their simulations that the control function estimator performs better than the 2SLS estimator, but their setting differs in several aspects from our study.[19] As part of the robustness

[19] For example, they only consider the control function estimator for regressions involving a binary outcome and do not consider non-linear trends in the endogenous variable.

checks, we perform a simulation study to understand better the properties of our control function approach when the instrument is weak.

We assume independence is likely to hold as a person's genome is a random (re)combination of their parent's genes conditional on ancestry. Whilst assortative mating and dynastic effects could lead to a relationship between the instrument and the outcome that does not operate through the offspring's testosterone level, we note that spousal correlations of genotypes for observable traits such as BMI, depression or even education are at best modest (Conley et al., 2016). We therefore consider it unlikely that there is non-negligible assortative mating on genetic variants related to (unobservable) testosterone levels in men. As part of our robustness checks, we control for parental education, which could mediate the relationship between parental genes and offspring earnings, and note our main findings remain unchanged. Using UKHLS data, Hughes and Kumari (2019) show that their PGS (constructed using the same GWAS by Ohlsson et al. (2011)) is not associated with various confounders, such as self-reported health and marital status. This partly addresses our concern regarding exclusion restrictions.

We further consulted the GWAS catalogue (Sollis et al., 2023) to examine whether any of the genetic variants used in the construction of our PGS has been identified as a correlate of other traits that might plausibly violate the exclusion restriction. The studies listed in the database suggested that one of the genetic variants (rs5934505) is correlated with creatinine (Sakaue et al., 2021), which in turn has previously been linked to labour market outcomes (Böckerman et al., 2017). Creatinine levels are related to muscle mass (Samoszuk et al., 2020), which suggests that the association between rs5934505 and creatinine could reflect an effect of testosterone on creatinine levels through increases in muscle mass. In this case, the exclusion restriction would not be violated. However, to rule out such concerns, we conduct additional sensitivity analyses in which we condition on creatinine, which is also measured at the time of the nurse visit. We also conduct an analysis in which we construct our PGS excluding the genetic variant rs5934505. The main substantive findings are robust to both adjustments and results can be found in Table A.3. Finally, to lend further support to our analytical approach we check for mutually independent distributions to confirm that our genetic variants are not in linkage disequilibrium, a core

assumption for MR estimators (VanderWeele et al. 2014; Sanderson et al. 2022).[20] Separately, we estimate Pearson correlations between each of the genetic variants used in the analysis and find these are very close to zero.

Violation of monotonicity would imply that for some individuals the genetic variants used to construct our polygenic score are predictive of lower (rather than higher) testosterone levels, which does not appear to be plausible given Figure 2.

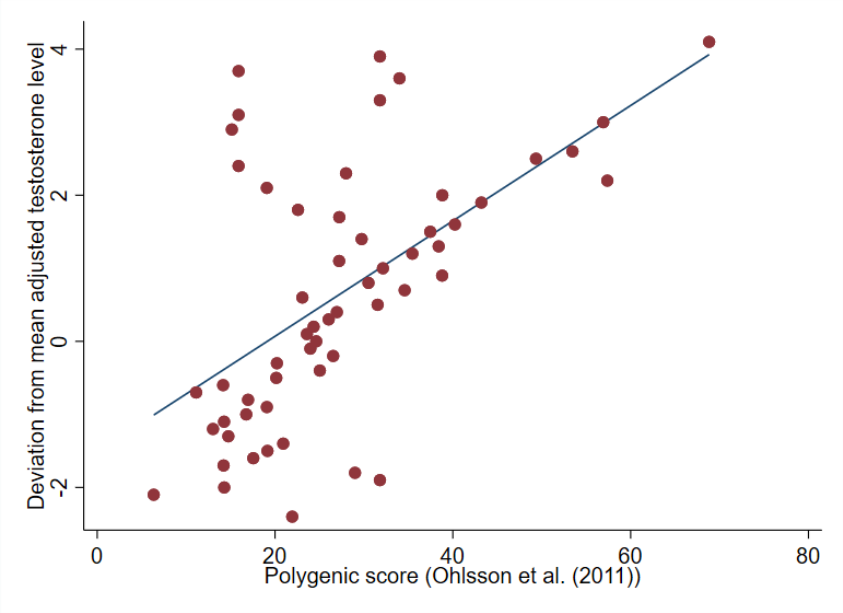
5 Results

5.1 First stage

Figure 2 depicts the relationship between the polygenic score (PGS) based on the GWAS conducted by Ohlsson et al. (2011) and deviations from average adjusted serum testosterone levels.[21] Higher values of the PGS are clearly associated with positive deviations from the reported average (adjusted) testosterone levels. It is unsurprising that the trend is approximately linear since the PGS is based on a linear combination of individual SNPs. However, we note that the Ohlsson et al. (2011) score is constructed using only three genetic variants, and consequently, the distribution of the PGS is multimodal (see Figure A.1 in the appendix). In addition to a linear trend, we therefore also consider including the PGS as a categorical variable using three categories (PGS<5, 5≥PGS<50 and PGS≥50), which reflects the limited variation in the score (see Figure A.1).[22]

[20] Linkage disequilibrium occurs when two or more genetic variants are correlated, which could imply a violation of the IV validity assumption.
[21] For Figure 2, we first calculate the deviation from the adjusted testosterone level, which ranges from -2.4 up to 4.2. We create 0.1 unit bins and calculate the mean PGS score for each bin separately.
[22] The PGS scores used to define the three categories reflect the 0-40th percentile, 41st-84th percentile and 85th percentile and above of the underlying distribution, respectively. These result in categories with 331, 986, and 304 observations for estimation purposes.

Figure 2: Deviation from average (adjusted) testosterone levels and the Ohlsson et al. (2011) polygenic score



Source: UKHLS, own calculations. *Notes*: Figure 2 depicts the deviation from the average level of adjusted testosterone within each PGS score bin. The scatterplot is overlayed with fitted values from a linear regression of the former variable on the latter.

Table 2 shows estimates from a regression of adjusted serum testosterone levels on the PGS. We present estimates for both scores based on Ohlsson et al. (2011) and Harrison et al. (2021). For the Ohlsson-PGS, we consider binary indicators for medium and high values of the PGS in addition to the linear functional form to account for the multimodal distribution of the score. Both PGS definitions have a strong positive effect on testosterone levels, i.e., individuals with a genetic predisposition towards higher testosterone levels have higher serum testosterone levels in the UKHLS sample. These estimates are highly statistically significant. The R^2 shows that the Ohlsson et al. (2011) PGS explains a larger share of the variance in testosterone levels than the modified Harrison et al. (2021) PGS. We see minimal differences in model fit between the two specifications for the Ohlsson et al. (2011) PGS. The model explains only around 5% of the variation in testosterone levels – testosterone levels vary substantially across individuals and are influenced by a range of factors, and we would therefore expect that even in a more saturated model, a large proportion of the variation in testosterone levels remains unexplained. We also estimate the squared semipartial correlation coefficient, which can be interpreted as a partial R^2. It suggests that the Ohlsson-PGS alone explains about 4.4% of the variation in testosterone levels.

Table 2: First stage regression

Dependent variable: Adjusted serum testosterone levels			
	PGS based on Ohlsson et al. (2011)		**PGS based on Harrison et al. (2021)**
	Linear PGS	**Categorical PGS**	**Linear PGS**
PGS	0.199***		0.141***
	(0.023)		(0.023)
Low PGS		*Baseline*	
Medium PGS		0.278***	
		(0.054)	
High PGS		0.606***	
		(0.071)	
N	**1,621**	**1,621**	**1,621**
R^2	0.050	0.051	0.028
Squared semipartial correlation	0.044		0.023
F statistic	74.45	38.10	38.08

Source: UKHLS, own calculations. *Notes*: The PGS was constructed based on the Ohlsson et al. (2011) GWAS. The table shows estimates of a linear regression controlling for the first ten genetic principal components and age group. Testosterone levels are adjusted for mean differences by time of the nurse visit and age group, as well as differences in standard deviations across age groups. Robust standard errors in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1

5.2 Effects of testosterone on earnings

Table 3 shows estimates of the effect of testosterone on earnings from our control function regressions using the Ohlsson-PGS as an instrumental variable for serum testosterone levels. Control function estimation allows us to model flexibly the functional form of testosterone in our second-stage regression. We consider five different specifications: (i) linear, (ii) quadratic, and (iii) cubic trends in testosterone; separate binary indicators for men with (iv) medium and high testosterone levels, as well as a combined binary variable for (v) medium/high vs. low testosterone levels.

The estimated residual from the first stage regression is negative in all specifications, which indicates that OLS estimates would suffer from a downward bias.[23] Our control function estimates

[23] For estimation purposes the PGS enters the first stage regression as a linear covariate.

indicate that higher testosterone levels lead to increased earnings. The estimated coefficients using a linear trend are large but not statistically significant, whereas all other specifications show highly significant effects of testosterone.

The categorical specification in Table 3 suggests that the difference in earnings between men with low and high testosterone levels is not significant. This is also observed in Figure A.2 in the online appendix, which shows an inverse U-shaped relationship between predicted earnings and testosterone levels using the estimated coefficients from a model including a quadratic trend in testosterone.[24] The quadratic trend implies an average marginal effect (AME) of testosterone of 0.065 when evaluated at the mean of the testosterone distribution. In contrast, the effect at the 25th and 75th percentile is 0.095 and 0.039, respectively, indicating that the marginal effect of testosterone on earnings is stronger for men with lower testosterone levels. For the following analysis, we use a quadratic trend as our preferred functional form for testosterone because a quadratic polynomial might capture the functional form of testosterone more flexibly than dummy variables for broad categories when we consider other outcomes as potential mechanisms.

[24] We obtain a very similar figure using the cubic trend (not shown, however available on request).

Table 3: Effect of testosterone on income

Dependent variable: Ln(total income)					
	Linear trend	**Quadratic trend**	**Cubic trend**	**Categorical**	**Binary**
T	0.057	0.065	0.055		
	(0.059)	(0.061)	(0.059)		
T^2		-0.022***	-0.031***		
		(0.008)	(0.010)		
T^3			0.005		
			(0.005)		
Low T				*Baseline*	-0.117***
					(0.036)
Medium T				0.104***	
				(0.037)	
High T				0.050	
				(0.070)	
Medium/High T					*Baseline*
First-stage residual	-0.055	-0.050	-0.055	-0.009	-0.028*
	(0.060)	(0.061)	(0.059)	(0.023)	(0.016)
N	1,621	1,621	1,621	1,621	1,621

Source: UKHLS, own calculations. *Notes*: The PGS was constructed based on the Ohlsson et al. (2011) GWAS. The table shows control function estimates using a linear regression model. The model controls for the first ten genetic principal components and age. Standard errors based on 1,000 bootstrap replications in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1

5.3 Effects of testosterone across the earnings distribution

Next, we consider whether our main findings vary across the earnings distribution. To do this, we estimate a distributional regression splitting our main sample in two: individuals belonging to the (bottom) top quartile of the earnings distribution are defined as being (low) high earners.[25] Table A.1 reports estimation results based on our preferred quadratic specification. In the case of low earners, we find the coefficient estimates on the first and second order term for testosterone are

[25] Distributional regression offers several advantages versus alternative estimators such conditional quantile regression given our research context, for example modelling the entire conditional distribution of testosterone (see, e.g., Koenker et al., 2013).

close to zero and insignificant. In contrast, for high earners, the coefficients are similar in magnitude and trend to our main findings in Table 3.[26]

5.4 Mechanisms

Next, we consider selection into occupations and individual characteristics as potential mechanisms for the positive effect of testosterone on earnings. Selection into occupation implies that testosterone levels influence men's choices of occupations and jobs, in which case the positive effect of testosterone on earnings might occur because men with higher testosterone levels choose better compensated jobs.[27] Alternatively, our findings may be driven by the effect of testosterone on individual characteristics, which in turn affect their performance and compensation.

These two pathways are closely related and cannot always be disentangled – e.g., testosterone might influence risk-taking behaviour, and men with high testosterone levels might systematically choose occupations in which risk-taking is rewarded (e.g., sales or self-employment). In the following, we show that there is limited selection into occupation. Consistent with this, we also demonstrate that most of the effect of testosterone on earnings occurs within occupational groups. We then consider several individual characteristics that have been associated with testosterone in previous studies and examine whether testosterone affects individual traits for men within the same occupational group.

5.4.1 Selection into jobs and occupations

We examine the effect of testosterone on occupational choice by estimating a multinomial logistic regression model using the first digit of the ISCO08 code as a categorical measure of occupation. As before, we control for the estimated residuals from the first stage regression of testosterone on the Ohlsson-PGS to address endogeneity. Table 4 shows the estimated coefficients on the quadratic trend in testosterone levels for all occupational groups. Group 1 ("Managers") serves as the reference group. We find a negative linear term for group 5 ("Service and sales workers") and

[26] We note that the AMEs at the 25th, mean and 75th percentile of adjusted testosterone distribution for both low and high earners, reported in Table A.2 are insignificant.
[27] Reasons for such selection into better compensated jobs could be status-seeking behaviour, higher willingness to take riskier jobs, or aptitude for specific skills and tasks that are highly valued by employers.

positive quadratic terms for groups 7 ("Craft and related trades workers") and 8 ("Plant and machine operators and assemblers"). While these estimates suggest that selection into occupation might be present, the pattern is not clear, and we note that most of the significant coefficients are only significant at the 10 percent level. The AMEs (see Table A.4) suggest that men with medium or high levels of testosterone are less likely to select into group 5 ("Service and sales workers") as opposed to group 1 ("Managers"). For all other groups, AMEs are small (<5 percentage points) and statistically insignificant.

We conduct several additional analyses to ensure that our conclusions are robust to the choice of the analytical model. We estimate an ordered logistic regression model, which assumes that the ISCO08 1-digit groups are ordered by occupational status. Then, we consider a binary distinction between Groups 1 and 2 (managers and professionals) and all other occupational groups. Finally, we repeat these analyses using the 5-group version of the NSSEC ("National Statistics Socio-economic Classification") as an alternative measure of occupational class. The results are similar to those presented in Table 4 – we find a significant quadratic term in testosterone in the ordered logistic regressions for both the ISCO08 and the NSSEC, yet the AMEs tend to be very small and statistically insignificant. Neither the coefficients (see Table A.5) nor the AMEs[28] are significant when using a binary classification. We do not find any significant effects when estimating the effect of testosterone on occupation of an individual's first job.[29]

[28] Available on request.
[29] Results are available on request.

Table 4: Effect of testosterone on current occupation

Dependent variable: 1-digit ISCO08 code of the current job									
ISCO group	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
T	*Baseline*	0.059	0.300	0.492	-0.932*	-1.103	0.059	0.535	-0.086
		(0.434)	(0.412)	(0.556)	(0.554)	(53.010)	(0.464)	(0.440)	(0.487)
T^2	*Baseline*	0.049	0.006	0.091	0.081	-0.089	0.106*	0.127**	0.056
		(0.071)	(0.066)	(0.089)	(0.080)	(273.00)	(0.063)	(0.060)	(0.068)
N					**1,621**				

Source: UKHLS, own calculations. Notes: The PGS was constructed based on the Ohlsson et al. (2011) GWAS. The table shows control function estimates using a multinomial logistic regression model. The model controls for the first ten genetic principal components and age. Standard errors based on 1,000 bootstrap replications in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1

We also consider whether testosterone affects the choice of jobs with specific characteristics within the broader occupational groups. We use outcome variables that measure total working hours, whether a respondent's job regularly requires working overtime, or whether a respondent's compensation involves a performance pay component and wage growth.[30] All regressions control for occupational group and marginal effects are reported at the mean. In this case, we find that higher testosterone levels only increase the probability of working overtime (see Table A.6).

Finally, we examine whether the effects of testosterone are driven by differences within or between occupations by re-estimating Table 3 with a set of occupation-fixed effects based on the first digit of the ISCO08 code. The estimates in Table 5 are qualitatively similar to the overall earnings effects reported in Table 3. A back-of-the-envelope calculation based on the estimates in the last column using a binary variable for medium/high testosterone levels suggests that variation in earnings between occupations explains less than 20% of the overall effect of testosterone on earnings. Calculating the average marginal effects for our preferred specification using a quadratic polynomial, we find that controlling for occupation does not affect our estimates of the effect of testosterone at the lower end of the distribution (0.095 without controls for occupation vs. 0.090 with controls at the 25th percentile), but the effect of testosterone at the mean and upper end of the

[30] Wage growth is measured by the change in hourly wages between waves B (2010-11) and L (2020-21) for individuals who report earnings at both survey waves.

distribution is slightly larger (0.065 without controls vs. 0.070 with controls at the mean; 0.039 vs. 0.052 at the 75th percentile).

In summary, we consistently find that testosterone has a significant impact on occupational choice, albeit the pattern is not very clear and the AMEs for most occupational groups are close to zero and statistically insignificant. Moreover, our results suggest that a large proportion of the effect of testosterone on earnings occurs within occupations. Taken together, this implies that selection into occupation plays at most a minor role in explaining the returns to testosterone.

Table 5: Effect of testosterone on income - within occupation

Dependent variable: Ln(total income)					
	Linear trend	**Quadratic trend**	**Cubic trend**	**Categorical**	**Binary**
T	0.065	0.070	0.066		
	(0.050)	(0.047)	(0.049)		
T^2		-0.015**	-0.018**		
		(0.007)	(0.009)		
T^3			0.002		
			(0.004)		
Low T				*Baseline*	-0.093***
					(0.032)
Medium T				0.089***	
				(0.033)	
High T				0.077	
				(0.060)	
Medium/High T					*Baseline*
First-stage residual	-0.067	-0.063	-0.0647	-0.022	-0.026*
	(0.051)	(0.049)	(0.0495)	(0.020)	(0.016)
N	**1,621**	**1,621**	**1,621**	**1,621**	**1,621**

Source: UKHLS, own calculations. Notes: The PGS was constructed based on the Ohlsson et al. (2011) GWAS. The table shows control function estimates using a linear regression model. The model controls for the first ten genetic principal components and age. We include fixed effects for the first digit of the ISCO code. Standard errors based on 1,000 bootstrap replications in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1

5.4.2 Individual characteristics

We explore a range of individual characteristics that might plausibly link testosterone and earnings. First, we analyse the effects of testosterone on educational attainment. Even though testosterone levels in our study were measured when most individuals have already completed FT education, the effects in our regression models are identified by variations in testosterone levels due to genetic differences. These genetic differences are stable over the life course; hence, they identify variations in testosterone levels that can be considered stable over time. However, it is important to bear in mind that the effect of genetic variants on phenotypes may change over the lifecycle; for example, certain (different) variants may code for testosterone early in life and so affect outcomes such as education and first job, which we are unable to capture.[31]

Table A.7 reports estimates from a control function ordered probit regression for educational attainment. We do not find any significant effects of testosterone on educational attainment. We also consider personality traits based on the Big-5 Inventory (Openness, Neuroticism, Extraversion, Conscientiousness, and Agreeableness; see Table A.8) or cognitive functioning (results available on request) as potential mechanisms, but we do not find any evidence that these individual traits are linked to testosterone levels.[32]

Evidence suggests testosterone is linked to individual risk-taking behaviour which may in turn influence labour market outcomes including earnings (Apicella et al., 2008; Coates and Herbert, 2008; Ronay and Von Hippel, 2010; Stenstrom et al., 2011). UKHLS measures self-reported individual risk attitude at wave 1 of the survey (approximately 1-2 years prior to the nurse visit) using the following survey instrument: "Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks"? Individuals answer on a 0-10 scale, with 0 corresponding to avoiding taking risks and 10 being fully prepared to take risks, respectively.[33] Table A.9 shows higher levels of risk-taking behaviour are positively associated with earnings. The magnitude of coefficient estimates is between 0.2 and 0.3 and generally

[31] The GWAS of Ohlsson et al. (2011) is based on Caucasian men of European descent.
[32] Cognitive functioning is measured in two ways. The first measures individuals' practical numerical knowledge and is a count measure based on the number of problems solved. The second is based on the cognitive testing measures developed for the US Health and Retirement Study, in which individuals are scored based on a number series test. Higher values reflect better fluid reasoning and the ability to solve novel problems. Further details regarding the cognitive measures collected in UKHLS can be found in McFall (2013).
[33] The risk-taking question is included as part of the self-completion survey.

trends upwards across risk categories, which corresponds to an increase in earnings of around 3% at the mean. Importantly, conditional on risk attitude, the qualitative nature of our findings with respect to testosterone remains unchanged.

Our analyses of potential mechanisms suggest that neither selection into occupational groups nor differences in individual characteristics seem to explain the testosterone wage premium. There are a few possible explanations for this: First, our sample size only allows us to consider relatively broad occupational groups. It might be possible that testosterone affects the selection of men into more narrowly defined occupational groups within the same 1-digit ISCO08 category. We also only observe a few selected job characteristics. It seems highly plausible, e.g., that variation in tasks and occupational skills, which may interact with a biological marker such as testosterone, is linked to differences in earnings. Further, testosterone has important "activating" effects, i.e., a person's behavioural response in specific situations (e.g., negotiations, confrontations) depends on their testosterone level. Such behavioural responses are plausibly linked to personality traits, but their transitory nature makes retrospective measurement in a survey challenging.

5.5 Sensitivity analysis

We first consider the sensitivity of our results to the restrictions imposed on our working sample. We then show that our results are unaffected by the transformation of testosterone, before discussing several checks of the IV assumptions and estimation procedures.

5.5.1 Sample restrictions

First, we extend our sample to include all individuals aged 16-64 at the time of their nurse visit. The first column of Table A.10 shows that the estimated regression coefficients assuming a quadratic functional form in the second stage are almost identical to those reported in Table 3.

Next, we include men working part-time and earning less than £900 per month. The second column of Table A.10 shows that in this case for the second stage, both the first and second order term for testosterone are significant at the 5% level; we note that the magnitude of the former is over twice that relative to the findings reported in Table 3. Importantly, the average marginal effect is positive and significant at the 25th, mean and 75th percentile of testosterone distribution (0.174, 0.145 and 0.120, respectively) and consistent with our main findings that

the effects are strongest in the lower part of the earnings distribution. We next relax the condition that the nurse visit took place between 09:00 and 20:00 to allow for the possibility of including, for example, shift workers. In this case, column 4 of Table A.10 shows the main findings and coefficient estimates remain essentially unchanged. Finally, we restrict the sample to men whose testosterone level falls between 9-25 nmol/l which is considered normal based on clinical guidelines (see Benzeval et al., 2014). Column 5 of Table A.10 shows that despite the magnitude of the second order polynomial term on testosterone being roughly twice the size versus that reported in Table 3, the qualitative nature of the findings remain unchanged.

The longitudinal dimension of UKHLS allows us to assess whether our main findings extend beyond the closest mainstage survey interview adjacent to the nurse visit. Conceptually, our instrument predicts the stable component of testosterone levels among working age men, which in turn influences earnings and this relationship should thus hold across the lifecycle. Table A.11 in the Appendix reports the likelihood of an individual belonging to the top 10% of the age-earnings distribution across survey waves 5-12 (2014-2021). The results show the qualitative nature of our main findings concerning the relationship between testosterone and earnings holds up to wave 8, approximately five years following the nurse visit. It is important to note that we do not account for labour market transitions, nor survey attrition and restrict attention to individuals who report earnings at each wave.

5.5.2 Transformation of testosterone

MR exploits genetic variants as instruments, which should be independent of age and time of day effect. We therefore re-estimate equation (2) using raw rather than standardised testosterone adjusted for individual's age and time of nurse visit. Table A.12 shows the results are largely consistent with our main findings, though we note the first order term is smaller in magnitude and significant.

5.5.3 IV assumptions

The MR literature has proposed several robustness checks to detect pleiotropic effects, which would violate the exclusion restriction. However, most of these methods (e.g., MR-Egger regression) are only applicable in settings with a large number of genetic variants (Bowden et

al., 2015).[34] Here, we combine only three genetic variants into a single score, and we therefore rely on more traditional approaches from the IV literature to assess the robustness of our findings. Separately, evidence suggests that the level of bias in the structural equation is directly related to the strength of the association between genetic variants, in our case, the PGS and the phenotype (Burgess et al., 2011). To minimise such bias, the authors suggest that researchers should use individual-level data where possible.

Following Hughes and Kumari (2019) we consider a range of variables which, based on existing evidence, could potentially confound the relationship between testosterone and earnings. Table A.15 in the Appendix reports findings from separate reduced form regression estimates of selected sociodemographic characteristics on the PGS. UKHLS contains detailed information regarding individual-level characteristics, including marital status, self-reported health, smoking behaviour and medication (for central nervous system and beta-blockers). We find no association between such variables and our PGS; though we cannot fully rule out that other dimensions of health might be affected.

Confounding may also arise due to genetic effects by ancestry not captured by principal components, for example, at the family level (Morris et al., 2020) due to environmentally mediated effects that induce correlation in parents' PGSs. Another possibility is single-trait or cross-trait assortative mating in the parents' generation. Data constraints imply we cannot rule out that family-level effects bias our results. However, UKHLS collects retrospective data on parental education and economic status. Including these additional characteristics in our regression allows us to control for the impact of parental genes that are mediated by parental outcomes which influence offspring labour earnings. Table A.16 shows that after controlling for parental education and economic status, the main findings remain unchanged; noting that parental education and maternal presence in the household are strongly correlated with male offspring labour market earnings conditional on testosterone level.

In Section 4.3, we highlighted that our main findings are robust to controlling for creatinine, which has been linked with labour market outcomes (Böckerman et al., 2017). Separately, the qualitative nature of our findings remains unchanged if we construct the PGS using only

[34] For these reasons, we do not report the findings based on MR-Egger-type tests, though results are available on request.

genetic variants associated with testosterone alone. We also tested for a statistical association between testosterone and creatinine to better understand the causal pathway between testosterone and earnings. Table A.13 reports estimates from a regression of creatinine on testosterone and finds no statistical association based on our main sample. Finally, we find no association between creatinine and our PGS conditional on testosterone (see Table A.14). These findings, combined with the results in Appendix Table A.3 suggest that, based on our sample data, (i) creatinine has an independent effect on individual earnings, and (ii) it is unlikely to be on the causal pathway between testosterone and earnings and, thus, does not violate the exclusion restriction.

To assess the sensitivity of our findings to weak instrument problems, we conduct a small simulation study to compare the performance of our CF estimator to a 2SLS estimator using the squared PGS as an additional instrument (see appendix B for a detailed description of the approach and the results). First, we find that the 95% confidence interval achieves nominal coverage for both estimators, regardless of the instrument's strength. As expected, the power to reject the null hypothesis and the confidence interval's size depends on the instrument's strength. We also find that our CF estimator substantially outperforms the alternative 2SLS estimator even for weak first stages.

5.5.4 Model specification

The CF approach and specification outlined in equations (1) and (2) implies a linear correlation between the first and second stage residual terms. However, additional endogeneity may arise for example due to omitted variables which have a quadratic effect on earnings. Table A.17 and Table A.18 report estimated coefficients and AMEs of the effect of testosterone on earnings (with and without controls for ISCO respectively) when we include both the first and second order polynomial of the residual in the second stage regression. First, we note that neither the first nor second order term is significant at conventional levels, suggesting no relevant endogeneity is detected. Whilst the absolute magnitude of the coefficient on the quadratic term for testosterone is twice as large as that reported in Table 3; the AME effects and qualitative nature of the findings mirror our main results. Taken together the analysis suggest endogeneity of this type is unlikely to bias our estimates and supports our choice of regression specification.

6 Discussion

Biological processes can play an important role in determining labour market outcomes (Böckerman et al., 2017; Gielen et al., 2016). We show that testosterone, a hormone that is linked to both physical development and individual behaviour, has a causal effect on British men's labour market earnings. We exploit variation in genetic markers of testosterone following a Mendelian Randomisation approach and estimate control function regressions, which allow us to flexibly control for the functional form of testosterone. Previous studies using a similar identification strategy only considered linear relationships between testosterone and earnings and found limited (Hughes and Kumari, 2019) or no effects (Harrison et al., 2021). We provide evidence that the relationship between earnings and testosterone appears to be non-linear. Consistent with Gielen et al. (2016)'s findings on prenatal testosterone exposure, we find positive returns to circulating testosterone for men—however, these effects appear to be limited to the top half of the earnings distribution. Importantly, we find that the marginal effects are highest among males with relatively low testosterone levels. We do not observe significant returns at higher levels of the male testosterone distribution. The returns to testosterone are economically important: our preferred regression specification suggests that moving from the 25th percentile of the testosterone distribution to the 30th percentile would lead to an increase in earnings by £277 or 9.7%, which is larger than Dolton and Sandi (2017)'s estimate of the returns to education for British men (6%).

Our analysis of potential mechanisms allows us to rule out several important pathways that have been proposed in the literature. First, our findings remained almost unchanged once we control for occupational group. Moreover, we find minimal evidence that testosterone is associated with selection into occupation; implying that our results instead reflect a testosterone wage premium which occurs *within* occupational groups. We also find no evidence of a link between testosterone and other job characteristics, e.g., working hours, performance pay, or wage growth. We can further rule out educational attainment, cognitive functioning, or Big 5 personality traits as relevant mechanisms, as they appear to be unrelated to testosterone in our analysis. However, the effect of genetic variants on testosterone may change over the lifecourse. Put another way, variants different to those used in this paper may be associated with testosterone at an earlier point in the lifecycle, which in turn influences socioeconomic outcomes, but our analysis cannot capture this.

In line with previous studies (Sapienza et al., 2009; Stenstrom et al., 2011), we find testosterone is associated with risk tolerance. However, controlling for risk tolerance does not seem to affect our estimates, which suggests that other mechanisms might be at play. For example, our data does not allow us to examine selection into occupations below the level of 1-digit ISCO88 codes. It is possible that men with higher testosterone levels select into higher paid occupations and jobs within the broader groups examined here. Similarly, it is possible that testosterone might affect selection into jobs involving specific tasks or requiring certain skills, which are highly valued by employers. Testosterone may also activate advantageous behavioural responses in certain situations, such as salary negotiations, that are not well captured in our data.

Our measure of testosterone further restricts our analysis. Here, we only consider total serum testosterone measured at a single point in time. Alternative measures, such as bioavailable or free testosterone, might more accurately capture the underlying biological processes (Harrison et al., 2021). Following the organizational-activational hypothesis, it also appears necessary to consider both the effects of prenatal exposure (Gielen et al., 2016) and the effects of circulating testosterone jointly. Replicating our results on a larger sample would allow a more in-depth consideration of, e.g., effect heterogeneity. Finally, our results only apply to men, since testosterone levels of most women in the UKHLS data are below the detectable threshold. It would be highly interesting to examine whether similar effects on earnings can be found for women.

Methodologically, we note that CF estimators impose additional structural assumptions on the first stage compared to 2SLS estimates (Wooldridge, 2015). This assumption plausibly holds here, since our first stage closely resembles the regression models used to construct our instrument, i.e., the polygenic score (Ohlsson et al., 2011). Assessing whether the usual IV assumptions hold appears to be more problematic since most diagnostics and tests proposed in the literature either focus on 2SLS estimation or only consider a single (often binary) endogenous variable.

Most genetic variants only predict very limited variation in phenotypes, and this is also the case here – our preferred first stage specification only explains around 5% of the variation in circulating testosterone. We conduct a simulation study to examine how weak instrument problems might affect our CF regressions. Similar to Burgess and Thompson (2012), we find

that our CF estimator outperforms an alternative 2SLS specification. However, we cannot entirely rule out that our results are affected by weak instrument problems. Replicating our analysis on a larger sample might further mitigate concerns about weak instruments. Considering the validity assumption, we provide some evidence that our instrument does not correlate with other known correlates of testosterone. Moreover, our polygenic score only relies on a small number of genetic variants, which do not seem to correlate with phenotypes unrelated to testosterone. In summary, we argue that the IV assumptions should plausibly hold. Yet, further guidance is urgently needed on how these assumptions should be assessed for control function estimators involving flexible specifications in the endogenous variable.

In conclusion, we show that there is a testosterone wage premium for British men. Whether this premium is warranted or represents discrimination will depend on the mechanisms linking testosterone to men's earnings. Our analysis rules out several plausible mechanisms proposed in the literature, such as education, occupation or personality traits. While we cannot identify the precise mechanism, it is noteworthy that this premium occurs primarily for higher earning men with relatively low testosterone levels, which contradicts some widely held stereotypes about “testosterone-fuelled” behaviour. Focusing on the lower end of the testosterone distribution might be a helpful starting point for future studies which seek to identify the relevant mechanisms.

References

Angrist, J., Kolesár, M., 2024. One instrument to rule them all: The bias and coverage of just-ID IV. J. Econom. 240, 105398. https://doi.org/10.1016/j.jeconom.2022.12.012

Apicella, C.L., Dreber, A., Campbell, B., Gray, P.B., Hoffman, M., Little, A.C., 2008. Testosterone and financial risk preferences. Evol. Hum. Behav. 29, 384–390. https://doi.org/10.1016/j.evolhumbehav.2008.07.001

Archer, J., 2006. Testosterone and human aggression: an evaluation of the challenge hypothesis. Relatsh. Brain Aggress. 30, 319–345. https://doi.org/10.1016/j.neubiorev.2004.12.007

Arnold, A.P., 2009. The organizational-activational hypothesis as the foundation for a unified theory of sexual differentiation of all mammalian tissues. Horm. Behav. 55, 570–578. https://doi.org/10.1016/j.yhbeh.2009.03.011

Benzeval, M., Davillas, A., Kumari, M., Lynn, P., 2014. Understanding Society: The UK Household Longitudinal Study. Biomarker User Guide and Glossary. University of Essex, Colchester.

Böckerman, P., Bryson, A., Viinikainen, J., Hakulinen, C., Pulkki-Råback, L., Raitakari, O., Pehkonen, J., 2017. Biomarkers and long-term labour market outcomes: The case of creatine. J. Econ. Behav. Organ. 142, 259–274. https://doi.org/10.1016/j.jebo.2017.08.003

Bowden, J., Davey Smith, G., Burgess, S., 2015. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int. J. Epidemiol. 44, 512–525. https://doi.org/10.1093/ije/dyv080

Brosnan, M., Gallop, V., Iftikhar, N., Keogh, E., 2011. Digit ratio (2D:4D), academic performance in computer science and computer-related anxiety. Digit Ratio 2D4D Individ. Differ. Res. 51, 371–375. https://doi.org/10.1016/j.paid.2010.07.009

Burgess, S., Thompson, S.G., 2015. Mendelian Randomization: Methods for Using Genetic Variants in Causal Estimation. Chapman and Hall/CRC, London, UK.

Burgess, S., Thompson, S.G., 2012. Improving bias and coverage in instrumental variable analysis with weak instruments for continuous and binary outcomes. Stat. Med. 31, 1582–1600. https://doi.org/10.1002/sim.4498

Burgess, S., Thompson, S.G., CRP CHD Genetics Collaboration, 2011. Avoiding bias from weak instruments in Mendelian randomization studies. Int. J. Epidemiol. 40, 755–764. https://doi.org/10.1093/ije/dyr036

Carré, J.M., McCormick, C.M., 2008. Aggressive behavior and change in salivary testosterone concentrations predict willingness to engage in a competitive task. Horm. Behav. 54, 403–409. https://doi.org/10.1016/j.yhbeh.2008.04.008

Chance, S.E., Brown, R.T., Dabbs, J.M., Casey, R., 2000. Testosterone, intelligence and behavior disorders in young boys. Personal. Individ. Differ. 28, 437–445. https://doi.org/10.1016/S0191-8869(99)00110-5

Choi, S.W., Mak, T.S.-H., O'Reilly, P.F., 2020. Tutorial: a guide to performing polygenic risk score analyses. Nat. Protoc. 15, 2759–2772. https://doi.org/10.1038/s41596-020-0353-1

Coates, J.M., Herbert, J., 2008. Endogenous steroids and financial risk taking on a London trading floor. Proc. Natl. Acad. Sci. 105, 6167. https://doi.org/10.1073/pnas.0704025105

Conley, D., Laidley, T., Belsky, D.W., Fletcher, J.M., Boardman, J.D., Domingue, B.W., 2016. Assortative mating and differential fertility by phenotype and genotype across the 20th century. Proc. Natl. Acad. Sci. 113, 6647–6652. https://doi.org/10.1073/pnas.1523592113

Dabbs, J.M., 1992. Testosterone and Occupational Achievement. Soc. Forces 70, 813–824. https://doi.org/10.2307/2579755
Dabbs Jr., J.M., de la Rue, D., Williams, P.M., 1990. Testosterone and occupational choice: Actors, ministers, and other men. J. Pers. Soc. Psychol. 59, 1261–1265. https://doi.org/10.1037/0022-3514.59.6.1261
Davies, N.M., Holmes, M.V., Davey Smith, G., 2018. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ 362, k601. https://doi.org/10.1136/bmj.k601
Dolton, P., Sandi, M., 2017. Returning to returns: Revisiting the British education evidence. Labour Econ. 48, 87–104. https://doi.org/10.1016/j.labeco.2017.07.001
Dreher, J.-C., Dunne, S., Pazderska, A., Frodl, T., Nolan, J.J., O'Doherty, J.P., 2016. Testosterone causes both prosocial and antisocial status-enhancing behaviors in human males. Proc. Natl. Acad. Sci. 113, 11633. https://doi.org/10.1073/pnas.1608085113
Eibich, P., Kanabar, R., Plum, A., Schmied, J., 2022. In and out of unemployment—Labour market transitions and the role of testosterone. Econ. Hum. Biol. 46, 101123. https://doi.org/10.1016/j.ehb.2022.101123
Geniole, S.N., Bird, B.M., Ruddick, E.L., Carré, J.M., 2017. Effects of competition outcome on testosterone concentrations in humans: An updated meta-analysis. Horm. Behav., Hormones and Human Competition 92, 37–50. https://doi.org/10.1016/j.yhbeh.2016.10.002
Gielen, A.C., Holmes, J., Myers, C., 2016. Prenatal Testosterone and the Earnings of Men and Women. J. Hum. Resour. 51, 30–61. https://doi.org/10.3368/jhr.51.1.30
Greene, F.J., Han, L., Martin, S., Zhang, S., Wittert, G., 2014. Testosterone is associated with self-employment among Australian men. Econ. Hum. Biol. 13, 76–84. https://doi.org/10.1016/j.ehb.2013.02.003
Hammermesh, D.S., Biddle, J.E., 1994. Beauty and the Labor Market. Am. Econ. Rev. 84, 1174–1194.
Harrison, S., Davies, N.M., Howe, L.D., Hughes, A., 2021. Testosterone and socioeconomic position: Mendelian randomization in 306,248 men and women in UK Biobank. Sci. Adv. 7, eabf8257. https://doi.org/10.1126/sciadv.abf8257
Hönekopp, J., Bartholdt, L., Beier, L., Liebert, A., 2007. Second to fourth digit length ratio (2D:4D) and adult sex hormone levels: New data and a meta-analytic review. Psychoneuroendocrinology 32, 313–321. https://doi.org/10.1016/j.psyneuen.2007.01.007
Hughes, A., Kumari, M., 2019. Testosterone, risk, and socioeconomic position in British men: Exploring causal directionality. Soc. Sci. Med. 220, 129–140. https://doi.org/10.1016/j.socscimed.2018.11.004
Kanabar, R., Mazur, A., Plum, A., Schmied, J., 2022. Correlates of testosterone change as men age. Aging Male 25, 29–40. https://doi.org/10.1080/13685538.2021.2023493
Keane, M., Neal, T., 2023. Instrument strength in IV estimation and inference: A guide to theory and practice. J. Econom. 235, 1625–1653. https://doi.org/10.1016/j.jeconom.2022.12.009
Koenker, R., Leorato, S., Peracchi, F., 2013. Distributional vs. Quantile Regression. https://doi.org/10.2139/ssrn.2368737
Kowal, M., Sorokowski, P., Żelaźniewicz, A., Nowak, J., Orzechowski, S., Żurek, G., Żurek, A., Juszkiewicz, A., Wojtycka, L., Sieniuć, W., Poniatowska, M., Tarnowska, K., Kowalska, K., Drabik, K., Łukaszek, P., Krawczyk, K., Stefaniak, T., Danek, N., 2020. No relationship between the digit ratios (2D:4D) and salivary testosterone

change: Study on men under an acute exercise. Sci. Rep. 10, 10068. https://doi.org/10.1038/s41598-020-66915-9
Kreuz, L.E., Rose, R.M., 1972. Assessment of aggressive behavior and plasma testosterone in a young criminal population. Psychosom. Med. 34, 321–332. https://doi.org/10.1097/00006842-197207000-00006
Lee, D.S., McCrary, J., Moreira, M.J., Porter, J., 2022. Valid t-Ratio Inference for IV. Am. Econ. Rev. 112, 3260–3290. https://doi.org/10.1257/aer.20211063
Mazur, A., Lamb, T.A., 1980. Testosterone, status, and mood in human males. Horm. Behav. 14, 236–246. https://doi.org/10.1016/0018-506X(80)90032-X
Mazur, A., Michalek, J., 1998. Marriage, Divorce, and Male Testosterone. Soc. Forces 77, 315–330. https://doi.org/10.1093/sf/77.1.315
McCaul, K.D., Gladue, B.A., Joppa, M., 1992. Winning, losing, mood, and testosterone. Horm. Behav. 26, 486–504. https://doi.org/10.1016/0018-506X(92)90016-O
Montiel Olea, J.L., Pflueger, C., 2013. A Robust Test for Weak Instruments. J. Bus. Econ. Stat. 31, 358–369. https://doi.org/10.1080/00401706.2013.806694
Neyse, L., Johannesson, M., Dreber, A., 2021. 2D:4D does not predict economic preferences: Evidence from a large, representative sample. J. Econ. Behav. Organ. 185, 390–401. https://doi.org/10.1016/j.jebo.2021.02.029
Nicolaou, N., Patel, P.C., Wolfe, M.T., 2017. Testosterone and Tendency to Engage in Self-Employment. Manag. Sci. 64, 1825–1841. https://doi.org/10.1287/mnsc.2016.2664
Nye, J.V.C., Bryukhanov, M., Kochergina, E., Orel, E., Polyachenko, S., Yudkevich, M., 2017. The effects of prenatal testosterone on wages: Evidence from Russia. Econ. Hum. Biol. 24, 43–60. https://doi.org/10.1016/j.ehb.2016.11.003
Nye, J.V.C., Orel, E., 2015. The influence of prenatal hormones on occupational choice: 2D:4D evidence from Moscow. Personal. Individ. Differ. 78, 39–42. https://doi.org/10.1016/j.paid.2015.01.016
Ohlsson, C., Wallaschofski, H., Lunetta, K.L., Stolk, L., Perry, J.R.B., Koster, A., Petersen, A.-K., Eriksson, J., Lehtimäki, T., Huhtaniemi, I.T., Hammond, G.L., Maggio, M., Coviello, A.D., EMAS Study Group, Ferrucci, L., Heier, M., Hofman, A., Holliday, K.L., Jansson, J.-O., Kähönen, M., Karasik, D., Karlsson, M.K., Kiel, D.P., Liu, Y., Ljunggren, Ö., Lorentzon, M., Lyytikäinen, L.-P., Meitinger, T., Mellström, D., Melzer, D., Miljkovic, I., Nauck, M., Nilsson, M., Penninx, B., Pye, S.R., Vasan, R.S., Reincke, M., Rivadeneira, F., Tajar, A., Teumer, A., Uitterlinden, A.G., Ulloor, J., Viikari, J., Völker, U., Völzke, H., Wichmann, H.E., Wu, T.-S., Zhuang, W.V., Ziv, E., Wu, F.C.W., Raitakari, O., Eriksson, A., Bidlingmaier, M., Harris, T.B., Murray, A., de Jong, F.H., Murabito, J.M., Bhasin, S., Vandenput, L., Haring, R., 2011. Genetic Determinants of Serum Testosterone Concentrations in Men. PLOS Genet. 7, e1002313. https://doi.org/10.1371/journal.pgen.1002313
ONS, 2024. Consumer price inflation time series (MM23) [WWW Document]. URL https://www.ons.gov.uk/economy/inflationandpriceindices/timeseries/l522/mm23 (accessed 9.26.24).
Ou, J., Wu, Y., Hu, Y., Gao, X., Li, H., Tobler, P.N., 2021. Testosterone reduces generosity through cortical and subcortical mechanisms. Proc. Natl. Acad. Sci. U. S. A. 118, e2021745118. https://doi.org/10.1073/pnas.2021745118
Phoenix, C.H., Goy, R.W., Gerall, A.A., Young, W.C., 1959. Organizing action of prenatally administered testosterone propionate on the tissues mediating mating behavior in the female guinea pig. Endocrinology 65, 369–382. https://doi.org/10.1210/endo-65-3-369
Sakaue, S., Kanai, M., Tanigawa, Y., Karjalainen, J., Kurki, M., Koshiba, S., Narita, A., Konuma, T., Yamamoto, Kenichi, Akiyama, M., Ishigaki, K., Suzuki, A., Suzuki, K.,

Obara, W., Yamaji, K., Takahashi, K., Asai, S., Takahashi, Y., Suzuki, T., Shinozaki, N., Yamaguchi, H., Minami, S., Murayama, S., Yoshimori, K., Nagayama, S., Obata, D., Higashiyama, M., Masumoto, A., Koretsune, Y., FinnGen, Ito, K., Terao, C., Yamauchi, T., Komuro, I., Kadowaki, T., Tamiya, G., Yamamoto, M., Nakamura, Y., Kubo, M., Murakami, Y., Yamamoto, Kazuhiko, Kamatani, Y., Palotie, A., Rivas, M.A., Daly, M.J., Matsuda, K., Okada, Y., 2021. A cross-population atlas of genetic associations for 220 human phenotypes. Nat. Genet. 53, 1415–1424. https://doi.org/10.1038/s41588-021-00931-x

Samoszuk, M., Morgentaler, A., de Groot, M., van Solinge, W., Li, Y., Adair, F., Hoefer, I., Haitjema, S., 2020. Association of low testosterone with changes in non-cardiovascular biomarkers in adult men. Int. J. Impot. Res. 32, 167–175. https://doi.org/10.1038/s41443-019-0112-4

Sapienza, P., Zingales, L., Maestripieri, D., 2009. Gender differences in financial risk aversion and career choices are affected by testosterone. Proc. Natl. Acad. Sci. 106, 15268. https://doi.org/10.1073/pnas.0907352106

Schaal, B., Tremblay, R.E., Soussignan, R., Susman, E.J., 1996. Male Testosterone Linked to High Social Dominance but Low Physical Aggression in Early Adolescence. J. Am. Acad. Child Adolesc. Psychiatry 35, 1322–1330. https://doi.org/10.1097/00004583-199610000-00019

Sollis, E., Mosaku, A., Abid, A., Buniello, A., Cerezo, M., Gil, L., Groza, T., Güneş, O., Hall, P., Hayhurst, J., Ibrahim, A., Ji, Y., John, S., Lewis, E., MacArthur, J.A.L., McMahon, A., Osumi-Sutherland, D., Panoutsopoulou, K., Pendlington, Z., Ramachandran, S., Stefancsik, R., Stewart, J., Whetzel, P., Wilson, R., Hindorff, L., Cunningham, F., Lambert, S.A., Inouye, M., Parkinson, H., Harris, L.W., 2023. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. Nucleic Acids Res. 51, D977–D985. https://doi.org/10.1093/nar/gkac1010

Stenstrom, E., Saad, G., Nepomuceno, M.V., Mendenhall, Z., 2011. Testosterone and domain-specific risk: Digit ratios (2D:4D and rel2) as predictors of recreational, financial, and social risk-taking behaviors. Digit Ratio 2D4D Individ. Differ. Res. 51, 412–416. https://doi.org/10.1016/j.paid.2010.07.003

Stock, J., Yogo, M., 2005. Testing for Weak Instruments in Linear IV Regression, in: Andrews, D.W.K. (Ed.), Identification and Inference for Econometric Models. Cambridge University Press, New York, pp. 80–108.

White, I.R., 2010. simsum: Analyses of simulation studies including Monte Carlo error. Stata J. 10, 369–385.

Wingfield, J.C., 2017. The challenge hypothesis: Where it began and relevance to humans. Horm. Behav., Hormones and Human Competition 92, 9–12. https://doi.org/10.1016/j.yhbeh.2016.11.008

Wooldridge, J.M., 2015. Control Function Methods in Applied Econometrics. J. Hum. Resour. 50, 420–445.

Zhang, K., Yang, X., Zhang, M., Wang, C., Fang, P., Xue, M., Zhao, J., Gao, X., Pan, R., Gong, P., 2020. Revisiting the relationships of 2D:4D with androgen receptor (AR) gene and current testosterone levels: Replication study and meta-analyses. J. Neurosci. Res. 98, 353–370. https://doi.org/10.1002/jnr.24502

Appendix – Tables

Table A.1: Effects of testosterone across the earnings distribution

Dependent variable: Ln(total income)		
	Low earners	**High earners**
T	0.009	0.015
	(0.055)	(0.052)
T^2	0.006	-0.024***
	(0.007)	(0.007)
First-stage residual	0.000	0.021
	(0.056)	(0.053)
N	1,621	1,621

Source: UKHLS, own calculations. *Notes*: The PGS was constructed based on the Ohlsson et al. (2011) GWAS. The table shows control function estimates using a linear regression model. The model controls for the first ten genetic principal components and age. Standard errors based on 1,000 bootstrap replications in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1

Table A.2: Effects of testosterone across the earnings distribution – marginal effects

Dependent variable: Ln(total income)		
	Low earners	**High earners**
25th percentile	0.000	0.048
	(0.056)	(0.052)
Mean	0.009	0.015
	(0.054)	(0.051)
75th percentile	0.016	-0.013
	(0.054)	(0.051)
N	1,621	1,621

Source: UKHLS, own calculations. *Notes*: The PGS was constructed based on the Ohlsson et al. (2011) GWAS. The table shows control function estimates using a linear regression model. The model controls for the first ten genetic principal components and age. Standard errors based on 1,000 bootstrap replications in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1

Table A.3: Effect of testosterone on income – testing plausibility of IV assumptions

Dependent variable: Ln(total income)		
	Excluding rs5934505	**Controlling for Creatinine**
T	0.143*	0.066
	(0.086)	(0.061)
T^2	-0.020**	-0.020**
	(0.008)	(0.008)
First-stage residual	-0.132	-0.053
	(0.088)	(0.062)
N	1,621	1,621

Source: UKHLS, own calculations. *Notes*: The PGS was constructed based on the Ohlsson et al. (2011) GWAS. The table shows control function estimates using a linear regression model. The model controls for the first ten genetic principal components and age. Standard errors based on 1,000 bootstrap replications in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1

Table A.4: Average marginal effects for the multinomial logistic regression model of 1-digit ISCO group on testosterone

Dependent variable: 1-digit ISCO08 code of the current job									
ISCO group	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
AME evaluated at…									
25th percentile	0.022	0.012	0.042	0.020	-0.104	-0.022	0.001	0.036	-0.006
	(0.043)	(0.037)	(0.028)	(0.022)	(0.065)	(0.052)	(0.038)	(0.027)	(0.034)
50th percentile	-0.011	0.000	0.030	0.024	-0.064*	-0.019	-0.000	0.051	-0.012
	(0.042)	(0.036)	(0.036)	(0.034)	(0.035)	(0.032)	(0.036)	(0.039)	(0.029)
75th percentile	-0.035	-0.008	0.015	0.029	-0.041**	-0.013	-0.000	0.069	-0.016
	(0.037)	(0.036)	(0.041)	(0.049)	(0.018)	(0.020)	(0.038)	(0.054)	(0.026)

Source: UKHLS, own calculations. Notes: The PGS was constructed based on the Ohlsson et al. (2011) GWAS. The table shows average marginal effects derived from control function estimates using a multinomial logistic regression model. The model controls for the first ten genetic principal components and age. Standard errors based on 1,000 bootstrap replications in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1

Table A.5: Alternative classifications of occupation

	ISCO		**NSSEC**	
	Ordered model	Binary split	Ordered model	Binary split
T	0.052	-0.091	0.036	-0.146
	(0.199)	(0.248)	(0.223)	(0.307)
T^2	0.061**	-0.060	0.085***	-0.041
	(0.029)	(0.040)	(0.033)	(0.043)
N	1,621	1,621	1,621	1,621

Source: UKHLS, own calculations. Notes: The PGS was constructed based on the Ohlsson et al. (2011) GWAS. The table shows control function estimates using an ordered logistic regression (columns 1 and 3) or a linear regression model (columns 2 and 4). The binary split for ISCO groups contrasts groups 1 and 2 ("Managers" and "Professionals") to the remaining 6 groups. The binary split for the NSSEC contrasts group 1 ("Management and Professional") to the remaining 4 groups. The model controls for the first ten genetic principal components and age. Standard errors based on 1,000 bootstrap replications in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1

Table A.6: Effect of testosterone on job characteristics (AME)

Dependent variable	*Total working hours*	*Job regularly requires working overtime*	*Performance pay component*	*Wage growth*
T	1.067	0.121**	-0.016	-0.080
	(0.971)	(0.053)	(0.057)	(0.109)
N	**1,619**	**1,590**	**1,119**	**591**

Source: UKHLS, own calculations. *Notes*: The PGS was constructed based on the Ohlsson et al. (2011) GWAS. The table shows control function estimates using a linear regression model. The model controls for the first ten genetic principal components and age. Standard errors based on 1,000 bootstrap replications in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1

Table A.7: Testosterone and educational attainment

	AME of testosterone evaluated at:		
Qualification level	**25th percentile**	**Mean**	**75th percentile**
Degree	-0.025	-0.040	-0.051
	(0.042)	(0.038)	(0.035)
Other higher degree	-0.003	-0.006	-0.009
	(0.004)	(0.005)	(0.007)
A-level etc.	0.003	0.003	0.001
	(0.006)	(0.003)	(0.003)
GCSE etc.	0.012	0.019	0.023
	(0.019)	(0.017)	(0.015)
Other qualification	0.009	0.015	0.021
	(0.014)	(0.014)	(0.016)
No qualification	0.005	0.009	0.014
	(0.008)	(0.009)	(0.013)
N		**1,621**	

Source: UKHLS, own calculations. Notes: The PGS was constructed based on the Ohlsson et al. (2011) GWAS. The table shows average marginal effects for the six categories of educational attainment estimated from a control function ordered probit regression. The model controls for the first ten genetic principal components and age. Standard errors based on 1,000 bootstrap replications in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1

Table A.8: Testosterone and personality traits

	AME of testosterone evaluated at:		
Big-5 Inventory	**25th percentile**	**Mean**	**75th percentile**
Openness	0.125	0.134	0.141
	(0.144)	(0.139)	(0.137)
Neuroticism	0.024	0.059	0.088
	(0.162)	(0.159)	(0.161)
Extraversion	-0.101	-0.058	-0.022
	(0.155)	(0.149)	(0.149)
Conscientiousness	-0.020	-0.027	-0.033
	(0.124)	(0.122)	(0.123)
Agreeableness	0.006	-0.004	-0.013
	(0.123)	(0.121)	(0.123)
N		**1,482**	

Source: UKHLS, own calculations. Notes: The PGS was constructed based on the Ohlsson et al. (2011) GWAS. The table shows average marginal effects for the personality traits measured based on the Big-5 Inventory estimated from a control function ordered probit regression. The model controls for the first ten genetic principal components and age. Standard errors based on 1,000 bootstrap replications in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1

Table A.9: Testosterone and risk-taking behaviour

Dependent variable: Ln(total income)	
Risk score	
0	*reference*
1	0.241*
	(0.137)
2	0.202*
	(0.116)
3	0.237**
	(0.11)
4	0.225**
	(0.112)
5	0.107
	(0.103)
6	0.268**
	(0.105)
7	0.291***
	(0.103)
8	0.345***
	(0.104)
9	0.365***
	(0.125)
10	0.258**
	(0.116)
T	0.099
	(0.064)
T^2	-0.030***
	(0.010)
First-stage residual	-0.077
	(0.065)
N	1,621

Source: UKHLS, own calculations. *Notes*: The PGS was constructed based on the Ohlsson et al. (2011) GWAS. The table shows control function estimates using a linear regression model. The model controls for the first ten genetic principal components and age. Standard errors based on 1,000 bootstrap replications in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1

Table A.10: Effect of testosterone on income – sample restrictions

Dependent variable: Ln(total income)				
	Sample: aged 16-64	**Part-time (<£900)**	**Nurse visit**	**Sample: testosterone 9-25 nmol/l**
T	0.056	0.145**	0.066	0.0855
	(0.058)	(0.074)	(0.061)	(0.088)
T^2	-0.020**	-0.021**	-0.018**	-0.046**
	(0.008)	(0.009)	(0.008)	(0.018)
First-stage residual	-0.045	-0.138*	-0.056	-0.048
	(0.059)	(0.077)	(0.062)	(0.091)
N	**1,692**	**1,748**	**1,654**	**1,386**

Source: UKHLS, own calculations. *Notes*: The PGS was constructed based on the Ohlsson et al. (2011) GWAS. The table shows control function estimates using a linear regression model. The model controls for the first ten genetic principal components and age. Standard errors based on 1,000 bootstrap replications in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1

Table A.11: Testosterone and belonging to the top 10% of the age-earnings distribution

Wave	*AME of testosterone*	**N**
5	0.034	1,337
	(0.042)	
6	0.031	1,221
	(0.044)	
7	0.029	1,140
	(0.049)	
8	0.052	1,069
	(0.045)	
9	-0.031	974
	(0.046)	
10	0.028	896
	(0.046)	
11	-0.005	835
	(0.046)	
12	0.059	750
	(0.055)	

Source: UKHLS, own calculations. Notes: The PGS was constructed based on the Ohlsson et al. (2011) GWAS. The model controls for the first ten genetic principal components and age. Standard errors based on 1,000 bootstrap replications in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1

Table A.12: Effect of testosterone on income – transformation of testosterone

Dependent variable: Ln(total income)	
	raw testosterone levels
T	0.031**
	(0.015)
T^2	< -0.001**
	(0.000)
First-stage residual	-0.0120
	(0.0121)
N	**1,692**

Source: UKHLS, own calculations. *Notes*: The PGS was constructed based on the Ohlsson et al. (2011) GWAS. The table shows control function estimates using a linear regression model. The model controls for the first ten genetic principal components and age. Standard errors based on 1,000 bootstrap replications in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1

Table A.13: The relationship between creatinine on testosterone

Dependent variable: creatinine	
T	-0.455
	(0.326)
N	**1,692**

Source: UKHLS, own calculations. *Notes*: Above shows a regression of testosterone on creatinine (also controlling for age and time of the day). Significance: *** p<0.01, ** p<0.05, * p<0.1

Table A.14: The relationship between creatinine on PGS and testosterone

Dependent variable: creatinine	
PGS (Ohlsson et al., 2011)	0.110
	(0.311)
T	-0.544*
	(0.329)
N	**1,621**

Source: UKHLS, own calculations. *Notes*: Above shows a regression of testosterone on creatinine (also controlling for age and time of the day). Significance: *** p<0.01, ** p<0.05, * p<0.1

Table A.15: Relationship between sociodemographic characteristics and PGS score

Outcome variable	**PGS score**	**N**
Medication: Beta blocker	-0.155	**1,692**
	(0.149)	
Medication: Central nervous system	0.071	**1,692**
	(0.085)	
Self-rated health (binary)	-0.109	**1,692**
	(0.072)	
Self-rated health (continuous)	0.0349	**1,692**
	(0.022)	
Married	0.029	**1,692**
	(0.050)	
Smoking	0.040	**954**
	(0.064)	

Source: UKHLS, own calculations. *Notes*: The PGS was constructed based on the Ohlsson et al. (2011) GWAS. The table shows control function estimates using a linear regression model. The model controls for the first ten genetic principal components and age. Standard errors based on 1,000 bootstrap replications in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1

Table A.16: Testosterone and parental education and economic status

Dependent variable: Ln(total income)	
Father working when 14	
Father working	*reference*
Father not working or deceased	-0.018
	(0.053)
Father not living with respondent so don't know or missing	0.102
	(0.078)
Mother working when 14	
Mother working	*reference*
Mother not working or deceased	-0.031
	(0.028)
Mother not living with respondent so don't know or missing	-0.288***
	(0.104)
Father's educational qualifications	
Did not go to school at all/left school with no qualifications or certificates	*reference*
Left school with some qualifications or certificates	0.062
	(0.038)
Gained further qualifications or certificates after leaving school	0.108***
	(0.035)
Gained a university degree or higher degree	0.236***
	(0.054)
Mother's educational qualifications	
Did not go to school at all/left school with no qualifications or certificates	*reference*
Left school with some qualifications or certificates	0.098***
	(0.038)
Gained further qualifications or certificates after leaving school	0.104**
	(0.043)
Gained a university degree or higher degree	0.173***
	(0.067)
T	0.059
	(0.056)
T^2	-0.021**
	(0.008)
First-stage residual	-0.048
	(0.057)
N	1,340

Source: UKHLS, own calculations. *Notes*: The PGS was constructed based on the Ohlsson et al. (2011) GWAS. The table shows control function estimates using a linear regression model. The model controls for the first ten genetic principal components and age. Standard errors based on 1,000 bootstrap replications in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1

Table A.17: Effect of testosterone and residuals on earnings

		AME of testosterone evaluated at:		
	Quadratic trend	**25th percentile**	**Mean**	**75th percentile**
T	0.0677	0.123*	0.0676	0.0208
	(0.0583)	(0.0699)	(0.0589)	(0.0673)
T^2	-0.0408			
	(0.0266)			
First order polynomial term in the first stage residual	-0.0527 (0.0597)			
Second order polynomial term in the first stage residual	0.0196 (0.0275)			
N	1,621	1,621	1,621	1,621

Source: UKHLS, own calculations. *Notes*: The PGS was constructed based on the Ohlsson et al. (2011) GWAS. The table shows control function estimates using a linear regression model. The model controls for the first ten genetic principal components and age. Standard errors based on 1,000 bootstrap replications in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1

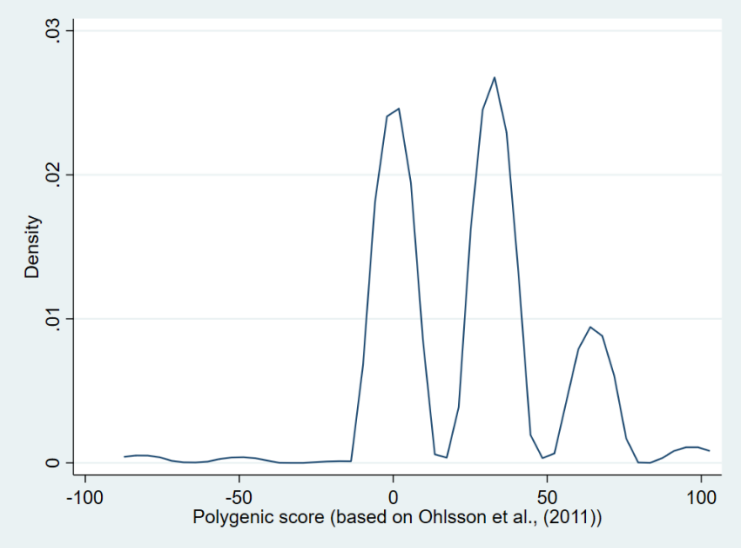
Table A.18: Effect of testosterone and residuals on earnings- within occupation

		AME of testosterone evaluated at:		
	Quadratic trend	**25th percentile**	**Mean**	**75th percentile**
T	0.0732	0.129**	0.0732	0.0260
	(0.0469)	(0.0606)	(0.0500)	(0.0565)
T^2	-0.0410*			
	(0.0233)			
First order polynomial term in the first stage residual	-0.0669 (0.0480)			
Second order polynomial term in the first stage residual	0.0279 (0.0236)			
N	1,621	1,621	1,621	1,621

Source: UKHLS, own calculations. Notes: The PGS was constructed based on the Ohlsson et al. (2011) GWAS. The table shows control function estimates using a linear regression model. The model controls for the first ten genetic principal components and age. We include fixed effects for the first digit of the ISCO code. Standard errors based on 1,000 bootstrap replications in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1

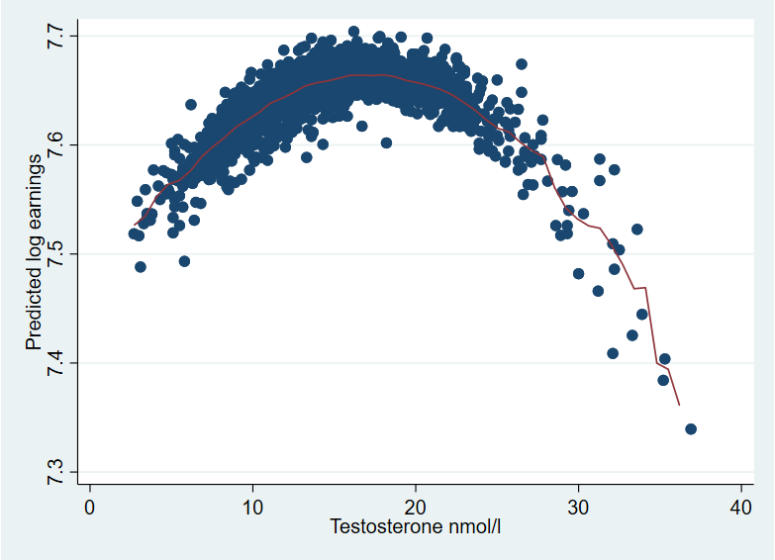
Appendix – Figures

Figure A.1: Distribution of the polygenic score



Source: UKHLS, own calculations. *Notes*: Figure A.1 depicts the distribution of polygenic scores based on the estimation sample.

Figure A.2: Predicted earnings and testosterone levels



Source: UKHLS, own calculations. *Notes*: Figure A.2 depicts the predicted relationship between earnings and testosterone level.

Appendix B: Simulation study on weak instruments in a control function regression

To assess how the strength of the instrument affects the performance of our control function estimator we undertake a simulation study. The data-generating process used for this simulation study closely follows the set-up in Keane and Neal (2023) with one important exception – our interest in the control function estimator is motivated by the ability to flexibly model the functional form of the endogenous variable in the structural equation, and we therefore include a quadratic trend in the endogenous variable in the second stage regression.

Here, $y \in \mathbb{R}$ is our outcome of interest and it is generated following the structural equation:

$$y_i = 0.3t_i + 0.03t_i^2 + u_i \qquad (1)$$

with $u_i \sim N(0,1)$. The endogenous variable $t \in \mathbb{R}$ is determined as a function of the instrumental variable $z \in \mathbb{R}$ based on the following first-stage equation:

$$t_i = \pi z_i + e_i, \qquad (2)$$

with $z_i \sim N(0,1)$, $e_i = \rho u_i + \sqrt{1-\rho^2}\,\eta_i$, and $\eta_i \sim N(0,1)$. In this setting, the endogeneity of t_i in eq. (1) stems from the correlation of the error terms of the structural equation, u_i, and the first-stage, e_i. The parameter ρ controls the degree of endogeneity. We set $\rho = 0.5$. The coefficient π in the first-stage allows us to control the size of the population F-statistic, $F^{pop} = N\pi^2$.

For our simulation, we generate a random sample of size $N = 1{,}000$ using random draws for z, u and η. We consider eight different values for π, corresponding to population F-statistics of 1, 3, 5, 10, 20, 50, 100 and 1000.[35] For each value of π, we then calculate e, z and y. We estimate the structural equation using both our control function estimator (with standard errors based on 200 bootstrap replications) as well as a 2SLS estimator with robust standard errors. For 2SLS estimation, we use z_i^2 as an additional instrument for t_i^2.

[35] These values are: 0.031622777, 0.054772256, 0.070710678, 0.1, 0.141421356, 0.223606798, 0.316227766 and 1.

We repeat this process using 1,000 randomly generated datasets. We evaluate the performance of both estimators by calculating summary statistics across these 1,000 simulations for each level of the population F-statistic using the Stata package 'simsum' (White, 2010).

The results are presented in Table B.1 below. First, we note that both estimators achieve nominal coverage of the 95% confidence interval for both the linear and quadratic term of the endogenous variable. However, the power to reject the null hypothesis using a 5% confidence level is generally poor for $F < 100$, and for the quadratic term the 2SLS estimator performs poorly even with $F = 1000$. As expected, bias of the point estimate and width of the 95% confidence interval decrease with increasing strength of the instrument. Notably, for $F \geq 3$, the control function estimator outperforms the 2SLS estimator for the linear term, and the control function estimator always performs better than the 2SLS estimator for the quadratic term of the endogenous variable.

While these results suggest that the bias and precision of our control function estimator depend on the strength of the instrument, they also show that even for weak instruments the nominal 95% confidence interval achieves correct coverage ensuring valid inference, and our control function estimator outperforms the 2SLS estimator in almost all scenarios considered here.

Table B.1*:* Comparison of estimators under different scenarios of instrument strength

	2SLS					Control Function				
Population F-statistic	Mean of point estimate	Relative bias	Mean width of 95% CI	% coverage of nominal 95% CI	% power of 5% level test	Mean of point estimate	Relative bias	Mean width of 95% CI	% coverage of nominal 95% CI	% power of 5% level test
Parameter: βt_i=0.3										
1	0.251	0.165	949.367	99.0	3.8	0.621	1.068	64.765	99.2	2.3
3	0.044	0.853	1525.510	98.7	4.6	0.191	0.362	42.157	98.7	4.4
5	0.233	0.223	676.003	98.8	6.2	0.242	0.192	31.175	98.1	7.9
10	0.209	0.302	742.052	98.5	9.5	0.243	0.189	16.517	97.0	16.0
20	0.249	0.169	174.480	98.2	13.1	0.274	0.087	3.816	96.6	29.8
50	0.229	0.237	453.794	97.7	32.5	0.293	0.022	0.615	96.0	52.6
100	0.286	0.047	3.945	96.9	64.4	0.298	0.007	0.408	95.7	78.5
1000	0.300	0.001	0.125	94.7	100.0	0.300	0.002	0.124	94.3	100.0
Parameter: βt_i^2=0.03										
1	0.558	17.594	1234.659	100.0	0.0	0.031	0.020	0.076	94.6	35.2
3	0.628	19.942	2418.448	100.0	0.0	0.031	0.020	0.076	95.1	36.0
5	1.912	62.733	5183.746	100.0	0.0	0.031	0.019	0.076	94.6	36.4
10	1.094	35.468	2409.435	100.0	0.0	0.031	0.018	0.076	94.6	36.4
20	-0.433	15.448	939.711	100.0	0.0	0.031	0.017	0.075	94.6	37.4
50	0.467	14.557	2578.053	99.8	0.2	0.030	0.014	0.073	93.9	38.2
100	-0.025	1.821	15.828	99.4	1.0	0.030	0.010	0.069	94.2	41.1
1000	0.028	0.052	0.089	95.9	24.5	0.030	0.008	0.038	95.0	85.4