

Height and Earnings: A Mendelian Randomization Analysis from China



Chen Zhu (CAU), Qihui Chen (CAU), Guoshu Kong (RUC),
Gang Chen (WeGene), Yingxiang Li (WeGene), Jun Wang (RUC)

- We collect both socioeconomic and genotyping data of 2,488 individuals from China.
- We analyze the impact of height on earnings by using two novel genetic instruments.
- 2SLS estimation results reveal *no causal* link between height and income.
- The “height premium” found from standard OLS regression is likely to pick up the impacts of other cognitive/noncognitive skills on earnings, rather than labor-market discrimination.



Collection of socioeconomic and genetic data using WeGene personal genetic testing kits and online surveys.

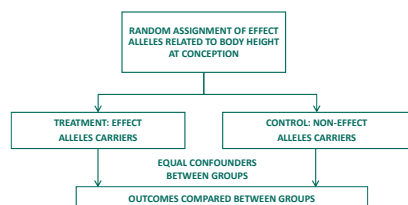
Introduction

> Taller people have been found to receive more years of education and earn more money in both developed and developing countries from observational studies (Case & Paxson, 2008; Magnusson & Rasmussen, 2006; Judge & Cable, 2006; Persico et al., 2004; Heineck, 2009; Gao & Smyth, 2010; LaFave & Thomas, 2017).

> Does body height causally affect socioeconomic status (SES)? Literature so far has only controlled for a limited number of socio-demographic/environmental factors and largely ignored the reverse causality of height, thus the estimated effect of height on SES from conventional observational designs may be vulnerable to the potentially endogenous height.

> In this study, we apply Mendelian randomization (MR, see Fig.1) to investigate the causal impact of height on earnings by using an original dataset recently collected in China containing prosperous genetic markers.

Fig.1 Design of MR



Mendelian Randomization

> In MR, genetic variants are randomly assigned at conception, and associations between genetic variants and outcome variables are not confounded by behavioral or environmental factors (Smith and Ebrahim, 2005; Smith and Hemani, 2014).

> Therefore, the concept of MR is analogous to randomized controlled trials (RCT) that are often infeasible or unethical and can help us make causal inferences about the effects of alcohol use/drinking behavior on individual health outcomes (von Hinke et al., 2016; Holmes et al., 2014; Lawlor et al., 2013; Yeung et al., 2012).

Contributions

> Our study is the first height-premium study that employs genetic IVs in China, if not the first in developing countries.

> The availability of two IVs allows us to formally test the validity of genetic IVs using the overidentification test.

> The demonstration of the validity of genetic IVs not only strengthens our own findings, but also lends support to previous MR findings by von Hinke et al. (2016), Tyrrell et al. (2016), and Böckerman et al. (2017).

> Our findings provide a deeper understanding of the “height premium” in China, which are valuable to achieving goals of reducing wage gap.

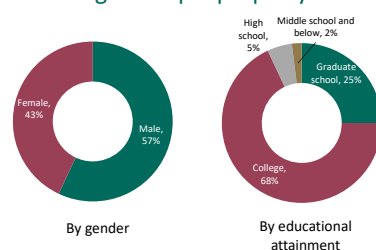
Data

> Our empirical analysis uses a novel dataset collected through collaboration with WeGene, a leading private genetic testing company in China that provides direct-to-consumer (DTC) genetic testing and personalized healthcare services.

> We conducted an online survey of WeGene customers. After providing informed consent, about 2,600 participants took our online survey between April 2018 and July 2018.

> Excluding individuals who were younger than the age of 16, older than 60, or still a student at the time of survey from the original dataset yields an analytical sample with 2,488 (see Fig.2).

Fig.2 Sample property



Genetic instruments

> Adult height is known as one of the highest heritable traits in human beings, determined approximately 80% by genetic factors and 20% by environmental factors (Yang et al., 2010; Jelenkovic et al., 2016).

> Thus, genetic variants related to height can naturally serve as unconfounded proxies for observed adult height (Böckerman et al., 2017; Hingorani and Humphries, 2005; Tyrrell et al., 2016; von Hinke et al., 2016).

> We construct a weighted polygenic score of height (PGScore_Height) based on 697 Single Nucleotide Polymorphisms (SNPs) that have been identified to be associated with adult height (Wood et al., 2014).

> A statistically significant and positive effect of genetic diversity has been found on human height (Koziel et al., 2011; Zhu et al., 2018). Thus, the second genetic IV is an index of individuals' ancestral diversity (AncesDiv), constructed based on the estimated composition of 42 ancestries from each respondent's genetic information.

Findings

> OLS results are consistent with previous observational studies conducted in China (Gao and Smyth, 2010; Yamamura et al., 2015), which suggest a statistically significant and positive association between one's body height and wage income: one additional cm in height is associated with 0.084%-1.34% increases in annual salary income (depending on different covariates included).

> In 2SLS regressions, the 1st stage F-statistics all range above 30, which exceed the traditional cut-off of 10 for weak instruments, and the Sargan statistics and *p*-values suggest no evidence that the genetic IVs are correlated with unobserved confounders, which provide strong supports to the validity of these two genetic IVs for body height in MR designs.

> We find *no* economically substantial causal effect of height on income, which suggests that discrimination from the workplace may not be a crucial component to wage disparity with respect to height. We interpret these results as empirical support of channels through which both cognitive and non-cognitive skills are correlated with height during individual development.



Follow us on WeChat
for more socioeconomic
studies! (In Chinese)

