Economics and Human Biology 24 (2017) 18–29

Contents lists available at ScienceDirect

Economics and Human Biology

journal homepage: www.elsevier.com/locate/ehb

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

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

[a] Turku School of Economics, Labour Institute for Economic Research, Helsinki, Finland
[b] Jyväskylä University School of Business and Economics, Jyväskylä, Finland
[c] School of Management, University of Tampere, Finland
[d] Unit of Psychology, University of Oulu, Oulu, Finland
[e] Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland
[f] Department of Clinical Chemistry, Fimlab Laboratories and School of Medicine, University of Tampere, Finland
[g] Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku and Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland

ARTICLE INFO

Article history:
Received 3 June 2016
Received in revised form 21 October 2016
Accepted 27 October 2016
Available online 8 November 2016

Keywords:
Height
Stature
Height premium
Earnings
Employment

ABSTRACT

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

1. Introduction

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

Most empirical studies treat height as an exogenous variable when examining the link between height and labor market outcomes.[3] However, there may be important confounders, such as socioeconomic background, early life conditions and parental investments, that influence both a person's height and subsequent labor market outcomes. It is challenging to adequately account for the combined effect of these factors when using the covariates that are available in observational studies. For example, parental investments are notoriously difficult to comprehensively and accurately measure.[4] Thus, the causal effect of height on earnings

* Corresponding author.
E-mail address: petri.bockerman@labour.fi (P. Böckerman).

[1] Hübler (2016) summarizes the relevant literature.

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

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

http://dx.doi.org/10.1016/j.ehb.2016.10.009
1570-677X/

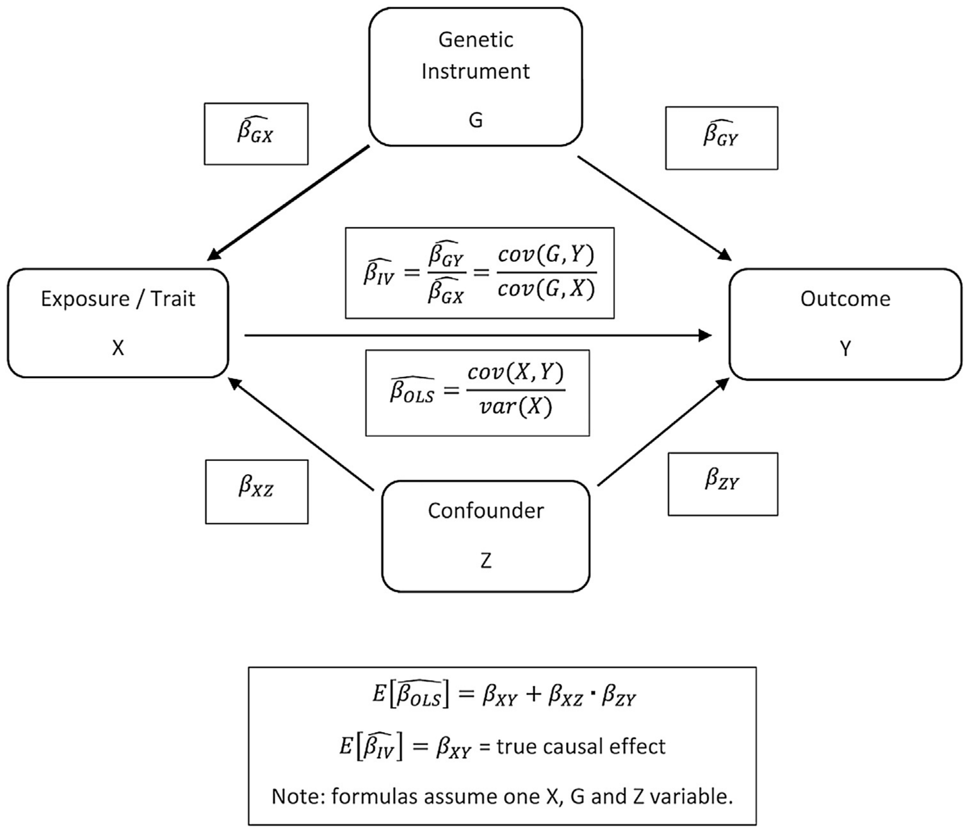

Fig. 1. Illustration of Mendelian randomization.

and labor market success largely remains an open question.

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

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

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

2. Methods

2.1. Mendelian randomization

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

Fig. 1 depicts the various effects and estimators in this setting. The IV or Wald estimator avoids the bias if the following conditions are fulfilled: (1) the genetic instrument (G) must correlate with the exposure variable (X), i.e. it must be informative; (2) the genetic instrument (G) must affect the outcome (Y) only through its effect on the exposure (X), i.e. the instrument must be exogenous; and (3) the instrument (G) and confounder (Z) must be independent, i.e.

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

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