Does education protect against depression? Evidence from the Young Finns Study using Mendelian randomization

Jutta Viinikainen[a,*], Alex Bryson[b,g], Petri Böckerman[a,g,h], Marko Elovainio[c], Niina Pitkänen[d], Laura Pulkki-Råback[c], Terho Lehtimäki[e], Olli Raitakari[f], Jaakko Pehkonen[a]

[a] University of Jyväskylä, Jyväskylä University School of Business and Economics, Jyväskylä, Finland
[b] University College London, London, United Kingdom
[c] Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Helsinki, Finland and National Institute for Health and Welfare, Helsinki, Finland
[d] Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland
[e] Department of Clinical Chemistry, Fimlab Laboratories and Finnish Cardiovascular Research Center Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland
[f] Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku and Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland
[g] IZA, Bonn, Germany
[h] Labour Institute for Economic Research, Helsinki, Finland

ARTICLE INFO

ABSTRACT

Using participants ($N = 1733$) drawn from the nationally representative longitudinal Young Finns Study (YFS) we estimate the effect of education on depressive symptoms. In 2007, when the participants were between 30 and 45 years old, they reported their depressive symptoms using a revised version of Beck's Depression Inventory. Education was measured using register information on the highest completed level of education in 2007, which was converted to years of education. To identify a causal relationship between education and depressive symptoms we use an instrumental variables approach (Mendelian randomization, MR) with a genetic risk score as an instrument for years of education. The genetic risk score was based on 74 genetic variants, which were associated with years of education in a genome-wide association study (GWAS). Because the genetic variants are randomly assigned at conception, they induce exogenous variation in years of education and thus identify a causal effect if the assumptions of the MR approach are met. In Ordinary Least Squares (OLS) estimation years of education in 2007 were negatively associated with depressive symptoms in 2007 (b = -0.027, 95% Confidence Interval (CI) = -0.040, -0.015). However, the results based on Mendelian randomization suggested that the effect is not causal (b = 0.017; 95% CI = -0.144, 0.178). This indicates that omitted variables correlated with education and depression may bias the linear regression coefficients and exogenous variation in education caused by differences in genetic make-up does not seem to protect against depressive symptoms.

1. Introduction

Approximately 6.8% of the world's population suffer from depression (Layard et al., 2013). The global direct and indirect costs of mental health conditions were about US$ 2.4 trillion in 2010 and the estimated loss in economic output due to mental health problems in 2010 was nearly US$ 16 trillion (Bloom et al., 2011). The high economic and societal costs of mental health conditions reinforce the need for policy measures that would prevent or alleviate these problems.

Several studies find an association between higher education and better health (see Madden, 2016) and improved mental health (e.g. Bauldry, 2015; Chevalier and Feinstein, 2007). Higher education can lead to better health outcomes in two ways. First, by enabling individuals to make lifestyle choices that are healthier in terms of diet and exercise. Second, higher education can improve individuals' marginal health returns from a given health input. For example, education may improve ability to comprehend advice from a doctor (Grossman, 1972; Grossman and Kaestner, 1997).

Education may also improve mental health indirectly through its impact on mediators such as higher earnings, thus providing financial

means for individuals to pursue better mental health (Cutler and Lleras-Muney, 2010) and reducing the risk of unemployment and work-related adverse life events which have been associated with mental distress (Audhoe et al., 2010). More educated individuals also tend to obtain jobs where control over job tasks is high and where there is a fair balance between effort and reward (Siegrist, 1996). These job characteristics help to reduce the mental strain associated with the job tasks (Karasek, 1979; Calnan et al., 2004). Additionally, education may improve the sense of mastery over life, improve one's relative position in society and help to create larger social networks, all of which are associated with better mental well-being (Dalgard et al., 2007; Marmot et al., 1997; Kawachi and Berkman, 2001).

Although education may improve mental health, it is an open question whether the correlation between education and depression is causal. Unobserved confounders influencing educational attainment and depressive symptoms, or reverse causality, may explain the observed correlation. To account for reverse causality and/or unobserved variables the literature has exploited data on twins, used statistical matching approaches and instrumental variables (IV) methods. The twin design accounts for shared family background and confounding genetic factors, which may otherwise cause omitted variable bias. Based on twin data, McFarland and Wagner (2015) and Mezuk et al. (2013) found that higher education was related to lower levels of depressive symptoms. Fujiwara and Kawachi (2009), who also use a twin design, however, did not find such a relationship. Bauldry (2015) uses propensity score matching to evaluate the link between education and depression and finds evidence for the protective effects of higher education against depressive symptoms. The method assumes that selection is based only on observed characteristics and thus omitted variable bias or the possibility of reverse causality cannot completely be ruled out.

Only a few studies aim to identify the causal effect of education on depression using an IV approach. Chevalier and Feinstein (2007) use longitudinal data form the British National Child Development Survey (NCDS) to estimate the connection between education and mental health. They use two different measures for mental well-being: a malaise score, which uses 24 items to assess depression and a self-reported measure of currently feeling sad or depressed. As instruments, Chevalier and Feinstein (2007) use teacher assessments concerning the benefits the child would get from post-compulsory schooling and the child's smoking behaviour at age 16. The latter is assumed to be a proxy for the discount rate (a higher discount rate implies lower investments in education). Based on the IV results education has a protective effect on mental health. The effect seems to be larger for women and for those with mid-level qualifications. Crespo et al. (2014) exploit schooling reforms in several European countries as instruments for educational attainment. They find that education has a large protective effect on mental health. However, other studies come to different conclusions. Kamhöfer et al. (2015) use variation in college availability and student loan regulations in Germany as instruments for higher education but do not find effects on mental health.

This paper examines whether higher education protects against depressive symptoms. Our contribution to the literature is to use a new instrument, a genetic-based risk score for years of education to identify the causal effect. Over the past decade this method, called Mendelian randomization (MR), has become an established tool for achieving causal inference in observational research (Pingault et al., 2018). Using rich, longitudinal population-based data which combines survey data with administrative register information on educational attainment and parental background we do not find causal support for the hypothesis that education would protect against depressive symptoms.

2. Methods

2.1. Data

The Cardiovascular Young Finns Study (YFS) is a longitudinal,

nationally representative study of 4320 individuals in six age cohorts (aged 3, 6, 9, 12, 15, 18) who were randomly chosen from five Finnish university hospital regions (Raitakari et al., 2008). In 1980 3596 subjects participated in the study and since then seven follow-up studies have been conducted.

In the YFS, depressive symptoms were evaluated in 2007 using a revised version of the Beck's Depression Inventory (BDI) (Beck and Steer, 1993; Elovainio et al., 2015; Rosenström et al., 2013). Depressive symptoms were measured using 21 items on a 5-point scale. The total depressive symptoms score is the average of all 21 items (Rosenström et al., 2012). The Beck's depressive symptoms scale is not a measure of clinically recognized psychiatric disorder nor does it indicate the chronicity or severity of depression (Elovainio et al., 2015).

To obtain information on the highest completed level of education, the YFS was linked to the Finnish Longitudinal Employer-Employee Data (FLEED) of Statistics Finland (SF) using unique personal identifiers. The matching is exact. The genetic risk score (GRS) we use as an instrument refers to years of education, so we used official guidelines from Statistics Finland to convert the education levels into years of education. For those who were still at school in 2007 (2.3% of $N = 1733$) years of education are based on the highest obtained degree.

The GRS is based on 74 single nucleotide polymorphisms (SNPs), which have been associated with the number of years of education of 293,723 individuals in a genome-wide association study (GWAS) (Okbay et al., 2016). As an instrument, the GRS has two key advantages over individual SNPs. First, the GRS accounts for more variation in years of education, which increases its statistical power. Second, the GRS reduces the risk that any individual SNP would bias the IV estimates via an alternative biological pathway (pleiotropy) (Palmer et al., 2012).

Genotyping was implemented by using the Illumina Bead Chip (Human 670K) and the genotypes were called using the Illumina clustering algorithm (Teo et al., 2007). SHAPEIT v1 and IMPUTE2 software (Delaneau et al., 2012) were used for genotype imputation and the 1000 Genomes Phase I Integrated Release Version 3 (March 2012 haplotypes) was used as a reference panel (Howie et al., 2009; 1000 Genomes Project Consortium, 2010). The GRS for educational attainment was based on 74 variants, which were associated with years of education (Okbay et al., 2016). Okbay et al. (2016) explored the mechanisms through which the candidate genes may affect education and found that many of these genes were related to brain development particularly during prenatal stages. The GRS was calculated as a sum of genotyped risk alleles or imputed allele dosages carried by an individual. The GRS was standardized to have a mean zero and standard deviation of one. Testing for the Hardy–Weinberg equilibrium (HWE) was conducted using the SNPTEST program (Marchini et al., 2007). Considering multiple testing, all 74 SNPs were in HWE ($p > 0.001$).

All models are adjusted for five residential regions in 2007, the birth year effects, gender and parental education. The data on parental education were drawn from Statistics Finland's Longitudinal Population Census (LPC) from the year 1980. It was linked to YFS-FLEED using unique identifiers. The indicator variable for high parental education equals one if at least one of the parents has obtained some university education.

Because of missing information on some variables the estimation sample is smaller than the total sample size. We tested the randomness of attrition with two-sample test of proportions (Table A1 in Supplementary Appendix). The results indicated that participants who were dropped from the analyses are more likely men and less educated. In terms of birth cohort or parental education there were no significant differences between the total YFS-FLEED-LPC sample and our final estimation sample.

2.2. Statistical methods

We first run Ordinary Least Squares (OLS) models to replicate