Author(s): Kari, Jaana T.; Viinikainen, Jutta; Böckerman, Petri; Tammelin, Tuija H.; Pitkänen, Niina; Lehtimäki, Terho; Pahkala, Katja; Hirvensalo, Mirja; Raitakari, Olli T.; Pehkonen, Jaakko

Title: Education Leads to a More Physically Active Lifestyle : Evidence Based on Mendelian Randomization

Year: 2020

Version: Accepted version (Final draft)

Please cite the original version:

Kari, J. T., Viinikainen, J., Böckerman, P., Tammelin, T. H., Pitkänen, N., Lehtimäki, T., Pahkala, K., Hirvensalo, M., Raitakari, O. T., & Pehkonen, J. (2020). Education Leads to a More Physically Active Lifestyle : Evidence Based on Mendelian Randomization. Scandinavian Journal of Medicine and Science in Sports, 30(7), 1194-1204. https://doi.org/10.1111/sms.13653

DR JAANA T KARI (Orcid ID : 0000-0001-5205-7031)

Article type : Original Article

Education Leads to a More Physically Active Lifestyle: Evidence Based on Mendelian Randomization

Running Head: education and physical activity

Kari JT*[1]., Viinikainen J.[1], Böckerman P.[1,2,3], Tammelin TH.[4], Pitkänen N.[5,6], Lehtimäki T[7]., Pahkala K[5,6,8], Hirvensalo, M[9]., Raitakari OT[5,6,10]., Pehkonen J[1].

[1] Jyväskylä University School of Business and Economics, University of Jyväskylä, Finland.
[2] Labour Institute for Economic Research, Helsinki Finland.
[3] IZA Institute of Labor Economics, Bonn, Germany.
[4] LIKES Research Centre for Physical Activity and Health, Jyväskylä, Finland.
[5] Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland.
[6] Centre for Population Health Research, University of Turku and Turku University Hospital, Turku, Finland.
[7] Department of Clinical Chemistry, Fimlab Laboratories and Faculty of Medicine and Health Technology, Finnish Cardiovascular Research Center - Tampere, Tampere University, Tampere, Finland.
[8] Paavo Nurmi Centre, Sports & Exercise Medicine Unit, Department of Physical Activity and Health, Turku, Finland.
[9] Faculty of Sport and Health Sciences, University of Jyväskylä, Finland
[10] Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland.

Corresponding author: Jaana T. Kari, DSc (Econ.), jaana.t.kari@jyu.fi, Jyväskylä University School of Business and Economics, P.O. Box 35, FI-40014 University of Jyväskylä, Jyväskylä, Finland. Tel. +358406234336.

Funding information

The Young Finns Study has been financially supported by the Academy of Finland: grant numbers 322098, 286284, 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117787 (Gendi), and 41071 (Skidi); the Social Insurance Institution of Finland; Competitive State Research Financing of the Expert Responsibility area of Kuopio, Tampere and Turku University Hospitals (grant number X51001); Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation for Cardiovascular Research; Finnish Cultural Foundation; The Sigrid Juselius Foundation; Tampere Tuberculosis Foundation; Emil Aaltonen Foundation; Yrjö Jahnsson Foundation; Signe and Ane Gyllenberg Foundation; Jenny and Antti Wihuri Foundation; Diabetes Research Foundation of Finnish Diabetes Association; EU Horizon 2020 (grant number 755320 for TAX-INOMISIS and grand number 848146 for TO AITION); European Research Council (grant number 742927 for MULTIEPIGEN project); and Tampere University Hospital Supporting Foundation.

Conflict of interest: The authors declare no conflicts of interest.

Abstract

Physical inactivity is a major health risk worldwide. Observational studies suggest that higher education is positively related to physical activity, but it is not clear whether this relationship constitutes a causal effect. Using participants (N = 1651) drawn from the Cardiovascular Risk in Young Finns Study linked to nationwide administrative data from Statistics Finland, this study examined whether educational attainment, measured by years of education, is related to adulthood physical activity in terms of overall physical activity, weekly hours of intensive activity, total steps per day, and aerobic steps per day. We employed ordinary least squares (OLS) models and extended the analysis using an instrumental variables approach (Mendelian randomization, MR) with a genetic risk score as an instrument for years of education. Based on the MR results, it was found that years of education is positively related to physical activity. On average, one additional year of education leads to a 0.62-unit higher overall physical activity ($p < 0.01$), 0.26 more hours of weekly intensive activity ($p < 0.05$), 560 more steps per day ($p < 0.10$), and 390 more aerobic steps per day ($p < 0.09$). The findings indicate that education may be a factor leading to higher leisure-time physical activity, and thus, promoting global health.

Keywords: Physical Activity, Education, Register-based data, Mendelian Randomization

1. Introduction

Highly educated individuals make healthier lifestyle choices; they are healthier and live longer. The literature has documented significant associations between education, general health,[1–6] and health behaviors.[5,7–14]

Physical activity is an important aspect of health behavior. Globally, approximately one-fourth of adults do not meet the recommended levels of daily physical activity for maintaining good health.[15] The importance of physical activity, as well as the detriments of physical inactivity, for health and well-being have been well documented. For instance, physical activity is related to a decreased risk of several chronic diseases,[16–18] and it may postpone the onset of dementia.[19] In contrast, physical inactivity has been identified as one of the leading risk factors for global mortality.[20] In 2013, the global economic burden of physical inactivity was estimated to range from INT$ 67.5 billion to INT$ 145.2 billion.[21] In general, the healthcare costs attributable to physical inactivity are estimated to range from 0.3% to 4.6% of the national healthcare expenditures.[22]

Empirical studies on the link between education and health behavioral outcomes, especially physical activity, have found that better academic achievement in adolescence and higher educational attainment in adulthood are related to higher physical activity. For example, Aaltonen et al.[7] showed that higher academic performance in adolescence was related to a higher frequency of self-reported leisure-time physical activity in young adulthood. In a recent study, Aaltonen et al.[23] also demonstrated that the association between self-reported leisure-time physical activity and academic achievement is partially explained by shared genetic background and family environment. Furthermore, Davies et al.[13] found that higher educational attainment was linked to higher levels of self-reported vigorous physical activity. Similar findings were documented by Park and Kang[24]: An increase in education induced individuals to exercise more regularly. Davies et al.[5], in turn, found only little evidence that higher educational attainment is related to physical activity. Based on accelerometer-measured physical activity, Kantomaa et al.[10] showed that a higher level of education was associated with a higher amount of moderate-to-vigorous physical activity, but at the same time, it was associated with lower levels of light-intensity activity and higher sedentary time. Based on a systematic literature review, Trost et al.[12] concluded that

education is a positive determinant of physical activity, whereas Bauman et al.[8] were more cautious about making explicit causal claims.

A growing body of population-based data includes both self-reported (questionnaires) and objective information (e.g., accelerometers or pedometers) of individuals' physical activity levels. However, most of the literature examining the relationship between education and physical activity is based on self-reported measures of physical activity. Furthermore, educational attainment is typically self-reported. To gain a better understanding of the links between education and physical activity, objective measures, along with self-reported information on physical activity, may be recommended. In this study, we examine whether educational attainment is related to adulthood physical activity. To address the existing research gap, we use both self-reported and device-based measures of physical activity, and we also use register-based information on educational attainment, which avoids biases resulting from self-reported measures. We first employ ordinary least squares (OLS) models, and extend the analysis using an instrumental variables approach known as Mendelian randomization (MR). MR is based on Mendel's law of segregation (the first law) and independent of assortment (the second law). The former states that alleles segregate randomly when they are passed from one generation to the next, and the latter states that the inheritance of one trait is independent of the inheritance of other traits.[25,26] This randomization causes exogenous variation in the exposure variable by nature, enabling causal identification.[25,27] We take advantage of the genetic risk score (GRS) for years of education as an instrument, which is based on 74 single-nucleotide polymorphisms (SNPs) related to educational attainment.[28] We hypothesize that higher education leads to a more physically active lifestyle.

2. Materials and methods

2.1 Study population

A study sample consisting of 1651 participants was drawn from three Finnish data sets: 1) the longitudinal Cardiovascular Risk in Young Finns Study (YFS), 2) the Finnish Longitudinal Employer-Employee Data (FLEED) of Statistics Finland, and 3) the Longitudinal Population Census (LPC) of Statistics Finland. (See Appendix 1 for the flow chart of the YFS-FLEED-LPC data). The three data sets were linked using personal identifiers.

The YFS was launched in the late 1970s to study cardiovascular risk in youth.[29] The first cross-sectional study was conducted in 1980, when 3596 participants in six age cohorts (aged 3–18 years) participated in the baseline study. The participants were randomly chosen boys and girls from the population registers of the five Finnish university hospital districts and their rural surroundings. Since 1980, seven follow ups have been conducted, with the latest one in 2011. FLEED is an annual panel covering the total working-age population of Finland. It records comprehensive register information on labor market outcomes, such as earnings and employment, and highest post-compulsory educational attainment. The data originate directly from tax and other administrative registers, and they are maintained by Statistics Finland. Register information on family background (parental education) was drawn from the LPC in 1980. The research protocol of the YFS has been approved by the ethics committees of the five universities, and all the participants have provided written informed consent.[29] The final linked YFS-FLEED-LPC analysis data have been approved for research purposes by Statistics Finland (Permission TK-53-673-13).

2.2 Self-reported and pedometer-measured physical activity

Information on physical activity in adulthood was drawn from the YFS in 2011, when the participants were aged 34–49 years. We formed four physical activity variables based on self-reports and daily steps monitored by pedometer: 1) overall leisure-time physical activity, 2) hours of weekly intensive (breathtaking and sweating) activity, 3) total steps per day, and 4) aerobic steps per day.

The first variable, overall physical activity, was based on five items concerning the intensity of physical activity, frequency of physical activity, hours per week spent on intensive physical activity, average duration of one physical activity session, and participation in organized physical activity (Appendix 2).[30] The responses were ranked on a 3-point scale, and the overall physical activity score was defined as a sum of the five items. Thus, the total score ranged from 5 (lowest physical activity level) to 15 (highest physical activity level) (Appendix 2).[30,31] The same overall physical activity information was collected for a subsample of YFS participants at age 15 years (n = 1761) and we used this information in a robustness check. Second, we used the hours of weekly intensive (breathtaking and sweating) activity (Appendix 2, Question 3)[30,31], as an alternative measure for self-reported leisure-time physical activity. In 2011, physical activity was also

measured with a pedometer (Omron Walking Style One HJ-152R-E) for seven consecutive days (see Hirvensalo et al.[32] for additional details of the measurement and classification protocol). The steps were expressed as total steps per day and aerobic steps per day. The total steps comprised every step that was taken during the day, including leisure-time and working time. The aerobic steps, in turn, were calculated automatically for continuous walking that lasted for more than 10 minutes without interruption at a pace of >60 steps/min. The steps measured using the Omron Walking Style pedometer were shown to be comparable to the steps measured with the ActiGraph accelerometer (GT1M), with a correlation coefficient of 0.94 ($p < 0.01$.[32] The use of pedometer-measured physical activity reduced the sample size to 1338 participants.

2.3 Register-based educational attainment

Information on the highest completed level of education in 2007 was drawn from the FLEED. The educational attainment levels were converted to years of education using Statistics Finland's official estimates for completing a specific degree as follows: upper secondary education and postsecondary nontertiary education = 12 years; short-cycle tertiary education = 14 years; bachelor or equivalent level = 16 years; master or equivalent level = 18 years; doctoral or equivalent level = 21 years.

2.4 Genetic risk score for years of education

The genetic risk score (GRS) for years of education was based on 74 single-nucleotide polymorphisms (SNPs), which were associated with years of education in a genome-wide association study (GWAS) consisting of 293,723 individuals (See Okbay et al.,[28] Supplementary information, pp. 12–13, for technical details of the selection of independent genome-wide significant SNPs).

Genotyping was implemented by using the Illumina Bead Chip (Human 670K) from 2442 YFS participants, including 546,677 SNPs. The genotypes were called using the Illumina clustering algorithm.[33] Quality control was performed using the Sanger genotyping QC pipeline, and individuals with possible relatedness were removed. Genotype imputation was conducted with the SHAPEIT v1 and IMPUTE 2 software,[34] and the 1000 Genomes Phase I Integrated Release version 3 (March 2012 haplotypes) was used as a reference panel.[35,36] The unweighted GRS we use is equal to the sum of the 74 alleles or imputed allele dosages that increase the probability that

one will complete a higher number of education years. In the weighted GRS (results presented in the appendixes), each risk allele or imputed allele dosage is multiplied by the effect sizes.[28] Both GRSs were standardized, with a mean zero and standard deviation of one. The Hardy–Weinberg equilibrium (HWE) test was performed using the SNPTEST program.[37] Considering multiple testing, all 74 SNPs were in HWE ($p > 0.001$). As an instrument, the GRS has two key advantages over individual SNPs. First, the GRS accounts for more variation in years of education; this increases its statistical power in instrumental variable estimation. Second, the use of the GRS reduces the risk of pleiotropy; that is, any individual SNP would bias the instrumental variables (IV) estimates via an alternative biological pathway.[38]

2.5 Confounding factors

The baseline models included only clearly exogenous and predetermined controls: gender (being female), birth year, and birth month. Thereafter, the models were also adjusted by family education. The indicator variable for high parental education equals one if at least one of the parents has obtained some university education by the year 1980. The inclusion of family education accounts for unobserved heterogeneity, such as innate ability and preferences, alleviating possible biases in the estimated correlation between education and physical activity.

2.6 Statistical analysis

To use our findings to replicate the standard observational studies of the literature, we first estimated OLS models. Because of potential confounders, together with reverse causation, the OLS regression coefficients may be biased.[39] Therefore, the analysis is extended with the IV method, that is, one-sample MR, in which the GRS for years of education was used as an instrument for educational attainment.[25,40]

The MR estimator avoids the bias related to the OLS estimator if the following four conditions are satisfied[25,27,41]: First, the genetic instrument is associated with the exposure of interest (*relevance assumption*). Second, the genetic instrument is independent of the factors that confound the association of the exposure (education) and outcome (physical activity); that is, the instrument is as good as randomly assigned (*independence assumption*). Third, the genetic instrument is exogenous; that is, the instrument is independent of the outcome, except possibly via its association with the exposure (*exclusion restriction assumption*). Finally, the instrument has a

monotonic effect on the exposure; that is, for a given change in the value of the genetic instrument, it cannot be that some individuals increase the treatment intensity while the others decrease the treatment intensity (*monotonicity assumption*).

The main concern related to MR is instrument validity. Potential threats to this validity are: 1) the frequency of the genetic variants varies in different subpopulations; 2) pleiotropy, that is, the genetic instrument affects the outcome variable either directly or through other pathways than the exposure variable; 3) the exposure variable is time-varying; 4) gene-environment interactions; 5) reverse causation; 6) the exposure variable is measured with error; and 7) other genetic markers in linkage disequilibrium with the one used in the analysis affect the outcome.[27,42–44]

We addressed these potential threats to identification in multiple ways. First, to minimize measurement error and problems related to time-varying exposure, information on individuals' educational attainment was drawn from the official registers in 2007, when the youngest YFS participants were 30 years old. Thus, the number of individuals who were still studying was very low (2.3%). Second, the Finnish population is ethnically homogenous reducing the possibility that the allele frequency will differ in different subgroups. We also tested whether the distribution of observable characteristics differs across the distribution of the GRS.[41] To account for genetic, (e.g., dynastic effects) and environmental effects, we included family controls—that is, parents' education—in our models. Third, to detect the potential alternative pathways through which SNPs in our GRS may affect physical activity, we used PhenoScanner, a publicly available database that provides summary results from GWAS.[45] Fourth, we ran Sargan's test, using 74 individual SNPs as instruments for education, to assess the validity of the overidentifying restrictions. Failure of the identification test would suggest that at least one of the genetic instruments is invalid. Fifth, we utilized a reduced-form model in which the outcome variable (physical activity) was explained by the GRS for education.[43] This approach does not rule out the possibility that the exclusion restriction assumption is violated, but it diminishes the potential biases resulting from time-varying exposure, gene-environment interactions, measurement error in the exposure variable, and reverse causation.[43] The reduced-form model identifies the effect of the exposure on the outcomes but not the quantitative size of the effect. Sixth, as an additional robustness check for instrument validity, we conducted a falsification test where leisure-time physical activity at age 15 was used as the dependent variable. Because adult educational attainment should not affect childhood

physical activity, a finding that adult education is not a predictor of child physical activity would be consistent with our identifying assumption.

3. Results

3.1 Descriptive evidence

The study sample consisted of 1651 individuals with information on the GRS, educational attainment, and leisure-time physical activity (overall physical activity and the hours of weekly intensive activity) and 1338 individuals with information on total steps per day and aerobic steps per day. The mean values of overall physical activity, hours spent in intensive activity, total steps, aerobic steps, and years of education were 9.07 (standard deviation [SD] 1.89), 3.50 (SD 1.34), 8024 (SD 3042), 1939 (SD 2102), and 13.88 (SD 2.68), respectively (Table 1). Women made up 56% of the sample, the average age in 2011 was 41 years, and 13% of the participants had at least one highly educated parent. According to the descriptive statistics, highly educated individuals (i.e. above the median years of education) tended to report higher leisure-time physical activity levels had more aerobic steps per day and fewer total steps per day compared with their less educated peers (i.e. below the median years of education). The difference in the GRS values between the more and less educated individuals was 0.20 units (p<0.001), supporting the relevance of the instrument (Table 1, Panel 1). Panel 2 of Table 1 compares individual differences by the instrumented value. Among high-GRS participants (above median GRS), the proportion of highly educated parents was higher compared with that of low-GRS participants. This pattern provides support for the importance of controlling for parental education.

To identify potential alternative pathways through which the SNPs in the education GRS may affect physical activity, we used PhenoScanner.[45] Of the 74 SNPs linked to education, some were also associated with obesity, height, waist-hip ratio, and body mass index. Therefore, the differences in these attributes between low- and high-GRS individuals were also examined (Table 1, Panel 2). However, we did not find differences in these attributes between these two groups.

<Table 1 >

3.2 OLS results

The OLS estimates show that the years of education is related to physical activity (Table 2). On average, one additional year of education is related to a 0.07-unit higher overall physical activity (b = 0.07; 95% confidence interval [CI] = 0.04 to 0.11), 0.04 more hours of intensive activity each week (b = 0.04; 95% CI = 0.02 to 0.07), 70 more aerobic steps per day (b = 70; 95% CI = 26 to 114), and 70 fewer total steps per day (b = –70; 95% CI = –132 to –8). The inclusion of family education as an additional control kept the education estimate largely intact (Table 2, Model 2). The results also did not change when the sample size was restricted for those for whom we had information on both self-reported and pedometer-measured physical activity (Appendix 3).

<Table 2 >

3.3 MR results

The MR results based on the unweighted GRS (Table 3, Model 1) imply that education increases physical activity. Appendix 4 presents the results for weighted GRS. On average, one additional year of education increases the overall physical activity score by 0.60 units (b = 0.62; 95% [CI] = 0.21 to 1.03), the amount of intensive activity by 0.26 hours per week (b = 0.62; 95% [CI] = 0.01 to 0.52), the amount of total steps per day by 560 steps (b = 260; 95% [CI] = -106 to 1225) and the amount of aerobic steps per day by 380 steps (b = 378; 95% [CI] = -50 to 824). When the models were adjusted by family education (Table 3, Model 2), the point-estimates suggested even stronger association between education and self-reported physical activity. However, in the case of pedometer-measured physical activity, the point estimates were no longer significant when the models were augmented with parental education.

<Table 3 >

The first-stage F-statistics in the baseline MR were 19.24 (self-reported) and 13.59 (pedometer-measured) (Table 3, Model 1), and the excluded instrument (i.e., the GRS for education) were related to education (b = 0.27; 95% CI = 0.15 to 0.40; b = 0.26; 95% CI = 0.12 to 0.39) in the first stage. This supports the relevance assumption of the MR method. A formal statistical test for instrument validity, Sargan's overidentification test, supported the null hypothesis that all 74 SNPs can be considered exogenous for overall physical activity (p < 0.20), intensive activity (p < 0.52), total steps per day (p < 0.35), and aerobic steps per day (p < 0.24) (Appendix 5). The results from

the reduced-form models (Table 4 and Appendix 6 for weighted GRS) are consistent with the MR results, lending further support to the conclusion that exogenous variation in education caused by genetic differences increases physical activity.

<Table 4 >

As a final robustness check, we conducted a falsification test, in which childhood self-reported leisure-time physical activity at age 15 was regressed on adulthood years of education. These results provided further support for the MR identification assumptions (Table 5). While the OLS results suggest a positive correlation between adulthood educational attainment and childhood physical activity, the association was eliminated using MR approach, as we expected. The finding that the OLS results imply a positive relationship between adulthood education and childhood physical activity suggests that there are unobserved confounders that bias the OLS results. This highlights the importance of using methods that address these biases.

<Table 5 >

4. Discussion

Using data drawn from a nationally representative longitudinal study combined with register information on post-compulsory education, this study utilized the OLS and MR approaches to identify the relationship between educational attainment and physical activity in adulthood. The results show that the years of education increases overall leisure-time physical activity, hours of intensive activity per week and aerobic steps per day. Concerning total steps per day, the results depend on the method used: The OLS results, which use the completed years of education as an explanatory variable, suggest a negative association between education and total steps, whereas with the MR method—which uses the GRS for years of education as an explanatory variable—suggests a positive association.

There may be several explanations for the findings. One potential explanation for the positive association between education and physical activity is that education is related to decision-making abilities, which may lead individuals to make healthier long-term decisions in their behavior,[46] for example, pursuing more physically active lifestyles. Individuals with higher/lower education may

also influence the health behaviors of others, and thus, one explanation may be the peer effects.[6] Education and physical activity may also be mediated by income: Higher education raises income levels, which in turn, provide more opportunities to invest in physical activity.[47] It has also been suggested that both shared genetic and family environment partly account for the association between leisure-time physical activity and academic performance.[23]

Our findings are consistent with prior observational studies, which have found positive relationships between physical activity and educational attainment.[7,8,10,11,13] For example, better academic performance in adolescence has been found to predict more frequent leisure-time physical activity in late adolescence and young adulthood.[7] In addition, post-compulsory education is shown to be positively related to physical activity.[10] In particular, moderate-to-vigorous physical activity is shown to be more common among highly educated individuals compared with those with lower levels of education. In line with our OLS results for total steps per day, the prior literature has also shown that higher educational level is related to lower amounts of light-intensity activity and greater sedentary time.[10] Davies et al.[13] also employed the MR method to investigate the links between education and physical activity. Using the same GRS for education as we did, they found a positive association between education and self-reported physical activity (moderate and vigorous). However, they did not use device-based measurements of physical activity, and thus, the studies complement each other.

There are issues that must be considered when interpreting the results. First, the physical activity measurements have limitations. The use of self-reported physical activity may cause measurement error bias,[48] but pedometer-measures also have limitations. Especially, pedometers do not provide information on non-ambulatory activities, such as gym workouts, swimming, cycling, or similar activities, nor they are not designed to accurately distinguish the intensity of physical activity. The modest correlations between the self-reported and pedometer-based measures (Appendix 7) suggest that each of the physical activity outcomes represents a different dimension of physical activity, which helps to understand the differences in the results. In addition, although this study included four measures of physical activity, we are not able to distinguish leisure-time, occupational, and commuting physical activity. Highly educated individuals, for example, may more likely have jobs in which they engage in less physical activity during working hours than individuals with lower education. However, highly educated individuals may engage in more

leisure-time physical activity than their less educated counterparts, as our results suggests. There may also be confounding factors (e.g., time preferences), which can arguably affect both education years and the level of physical activity. However, instrumental variables approach, and especially the use of genetic variants as determinants of educational attainment, should not be influenced by confounding or attenuation.[49]

Second, there may also be gender differences in the links between education and physical activity. Based on the correlation coefficients (Appendix 7), education was negatively associated with total steps per day among men, but the association was non-existing among women. Because of the small sample size, we are not able to estimate MR models separately for men and women. Using a larger sample size including both self-reported and pedometer-measured physical activity, future studies could shed more light on these potential gender differences.

Third, the MR approach identifies causal effect only if the instrument is valid. We tested and found support for instrument validity, but in the MR setting, it is impossible to prove the null hypothesis of instrument validity. If, for example, the genetic variants are pleiotropic, the MR results may be biased. We also tested the instrument validity with a falsification test. We are aware that years of education in adulthood may be associated with higher leisure-time physical activity in adolescence. This is possible if the same SNPs related to years of completed education are also associated with academic achievement in childhood and youth, which is further related to adolescent physical activity, as the earlier studies have suggested.[7] However, our falsification test results with the MR method showed that years of education in adulthood were not associated with adolescent physical activity. (The point estimates even turned negative). Thus, our results do not support this possibility.

Fourth, according to Table 1, the completed years of education and the GRS for education differed according to family education. Thus, to capture the genetic and environmental transmission of education,[50,51] the models were adjusted for parental education. Parents influence their children's educational outcomes not just by transferring their genes to the children but also by influencing their educational pathways directly, for example, by buying homes in the areas with better schools or providing a stimulating environment. Typically, in the genetic literature, family background is taken into account within-family-methods, which utilize information on sibling or parental

genotype. Unfortunately, such information was not available in our data. However, previous economic literature has viewed family education as a relevant control that may not only capture the genetic transmission but also the environmental transmission of traits.[50,52] Lastly, the local average treatment effects (LATEs), identified with the MR method, capture the average effect of education on physical activity among compliers, that is, among those whose years of education is increased via the impact of the 74 SNPs that comprise the instrument in MR. The variation in education due to other factors may lead to different conclusions.

Education is a key component of human capital. In addition to the positive economic consequences of higher education, such as better employment prospects, higher earnings, and economic growth,[53,54] our findings suggest that education may also lead to a more physically active lifestyle. The MR results imply that one additional year of education increases the level of overall leisure-time physical activity by about one unit, intensive activity per week by about 20 minutes, the amount of total steps by about 500 steps, and the amount of aerobic steps by about 400 steps per day. According to the self-reported questionnaire (Appendix 2), a one-unit increase in overall physical activity can be reached if, for example, one of the following alternatives occurs: 1) the frequency of intensive physical activity increases from "once a month or more" to "once a week," 2) the amount of weekly intensive activity increases from "1 hour a week" to "2–3 hours a week," or 3) the duration of physical activity sessions increases from "less than 20 minutes" to "20–40 minutes."

5. Perspective

This study investigated the relationship between educational attainment and physical activity in adulthood. Compared with previous observational studies suggesting an association between education and physical activity, this study corroborates the association by using OLS and MR estimation methods and including self-reported and pedometer-measured physical activity. From the public health perspective, our findings are twofold. First, our results show that the benefits of education are not only confined to economic outcomes, such as higher earnings and stronger labor market attachment, but also, they may cover additional domains like health behaviors (i.e. physical activity). Consequently, education may have positive externalities that extend beyond economic outcomes, increasing education's societal returns. Second, the finding that education is positively

related to physical activity may be an important link modifying the risk of chronic diseases during the life course, and it may serve as a partial explanation for the higher rates of morbidity and mortality among less educated individuals. From the policy perspective, the finding that education is related to different dimensions of physical activity can aid health promoters in implementing efficient tools for increasing physical activity, and thus promoting global health, among individuals from different socioeconomic backgrounds.

References

1. Böckerman P, Viinikainen J, Pulkki-Råback L, et al. Does higher education protect against obesity? Evidence using Mendelian randomization. *Prev Med (Baltim)*. 2017;101:195-198. doi:10.1016/j.ypmed.2017.06.015
2. Grossman M. On the concept of health capital and the demand for health. *J Polit Econ*. 1972;80(2):223-255. doi:10.7312/gros17812-004
3. Leino M, Raitakari OT, Porkka KV, Taimela S, Viikari. Jorma SA. Associations of education with cardiovascular risk factors in young adults: The Cardiovascular Risk in Young Finns Study. *Int J Epidemiol*. 1999;28:667-675.
4. Marioni RE, Ritchie SJ, Joshi PK, et al. Genetic variants linked to education predict longevity. *Proc Natl Acad Sci U S A*. 2016;113(47):13366-13371. doi:10.1073/pnas.1605334113
5. Davies NM, Dickson M, Smith GD, Van Den Berg GJ, Windmeijer F. The causal effects of education on health outcomes in the UK Biobank. *Nat Hum Behav*. 2018;2(2):117-125. doi:10.1038/s41562-017-0279-y
6. Brunello, G; Fort, M; Schneeweis, N; Winter-Ebmer R. The causal effect of education on health: what is the role of health behaviors? *Health Econ*. 2016;25(3):314-336. doi:10.1002/hec.3141
7. Aaltonen S, Latvala A, Rose RJ, Kujala UM, Kaprio J, Silventoinen K. Leisure-time physical activity and academic performance: Cross-lagged associations from adolescence to young adulthood. *Sci Rep*. 2016;6. doi:10.1038/srep39215
8. Bauman AE, Reis RS, Sallis JF, et al. Correlates of physical activity: Why are some people physically active and others not? *Lancet*. 2012;380(9838):258-271. doi:10.1016/S0140-6736(12)60735-1
9. Gage SH, Bowden J, Smith GD, Munafo MR. Investigating causality in associations between education and smoking: A two-sample Mendelian randomization study. *Int J Epidemiol*. 2018;47(4):1131-1140. doi:10.1093/ije/dyy131

10. Kantomaa MT, Tikanmäki M, Kankaanpää A, et al. Accelerometer-measured physical activity and sedentary time differ according to education level in young adults. *PLoS One*. 2016;11(7). doi:10.1371/journal.pone.0158902

11. Mäkinen TE, Sippola R, Borodulin K, et al. Explaining educational differences in leisure-time physical activity in Europe: The contribution of work-related factors. *Scand J Med Sci Sport*. 2012;22(3):439-447. doi:10.1111/j.1600-0838.2010.01234.x

12. Trost SG, Owen N, Bauman AE, Sallis JF, Brown W. Correlates of adults' participation in physical activity: review and update. *Med Sci Sport Exerc*. 1996;34(12). doi:10.1249/01.MSS.0000038974.76900.92

13. Davies NM, Hill WD, Anderson EL, Sanderson E, Deary IJ, Smith GD. Multivariable two-sample mendelian randomization estimates of the effects of intelligence and education on health. *Elife*. 2019;8:1-22. doi:10.7554/eLife.43990

14. Cowell AJ. The relationship between education and health behavior: Some empirical evidence. *Health Econ*. 2006;15(2):125-146. doi:10.1002/hec.1019

15. Guthold R, Stevens GA, Riley LM, Bull FC. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1·9 million participants. *Lancet Glob Heal*. 2018;6(10):e1077-e1086. doi:10.1016/S2214-109X(18)30357-7

16. Lee IM, Shiroma EJ, Lobelo F, et al. Effect of physical inactivity on major non-communicable diseases worldwide: An analysis of burden of disease and life expectancy. *Lancet*. 2012;380(9838):219-229. doi:10.1016/S0140-6736(12)61031-9

17. 2018 Physical Activity Guidelines Advisory Committee. *2018 Physical Activity Guidelines Advisory Committee Scientific Report*. Washington, DC; 2018.

18. Reiner M, Niermann C, Jekauc D, Woll A. Long-term health benefits of physical activity – A systematic review of longitudinal studies. *BMC Public Health*. 2013;13(1). doi:10.1186/1471-2458-13-813

19. Carvalho A, Rea IM, Parimon T, Cusack BJ. Physical activity and cognitive function in individuals over 60 years of age: A systematic review. *Clin Interv Aging*. 2014;9:661-682.

doi:10.2147/CIA.S55520

20. WHO. *Global Status Report on Noncommunicable Diseases*.; 2014. http://apps.who.int/iris/bitstream/10665/148114/1/9789241564854_eng.pdf?ua=1.

21. Ding D, Lawson KD, Kolbe-Alexander TL, et al. The economic burden of physical inactivity: A global analysis of major non-communicable diseases. *Lancet*. 2016;388(10051):1311-1324. doi:10.1016/S0140-6736(16)30383-X

22. Ding D, Kolbe-Alexander T, Nguyen B, Katzmarzyk PT, Pratt M, Lawson KD. The economic burden of physical inactivity: A systematic review and critical appraisal. *Br J Sports Med*. 2017;51(19):1392-1409. doi:10.1136/bjsports-2016-097385

23. Aaltonen S, Latvala A, Jelenkovic A, et al. Physical activity and academic performance: genetic and environmental associations. *Med Sci Sports Exerc*. 2020;52(2):381-390. doi:10.1249/MSS.0000000000002124

24. Park C, Kang C. Does education induce healthy lifestyle? *J Health Econ*. 2008;27(6):1516-1531. doi:10.1016/j.jhealeco.2008.07.005

25. Lawlor DA, Harbord RM, Sterne JAC, Timpson N, Smith GD. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Stat Med*. 2008;27(8):1133-1163. doi:10.1002/sim.3034

26. Mendel G. Experiments in plant hybridization. 1865. http://www.mendelweb.org/archive/Mendel.Experiments.txt.

27. Smith GD, Hemani G. Mendelian randomization: Geneticanchorsfor causal inference in epidemiological studies. *Hum Mol Genet*. 2014;23(R1). doi:10.1093/hmg/ddu328

28. Okbay A, Beauchamp JP, Alan Fontana M, et al. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*. 2016;533. doi:10.1038/nature17671

29. Raitakari OT, Juonala M, Rönnemaa T, et al. Cohort profile: The cardiovascular risk in young Finns study. *Int J Epidemiol*. 2008. doi:10.1093/ije/dym225

30. Hirvensalo M, Magnussen CG, Yang X, et al. Convergent validity of a physical activity questionnaire against objectively measured physical activity in adults: The Cardiovascular Risk in Young Finns Study. *Adv Phys Educ*. 2017;07(04):457-472.

doi:10.4236/ape.2017.74038

31. Telama R, Yang X, Leskinen E, et al. Tracking of physical activity from early childhood through youth into adulthood. *Med Sci Sports Exerc*. 2014. doi:10.1249/MSS.0000000000000181

32. Hirvensalo M, Telama R, Schmidt MD, et al. Daily steps among Finnish adults: Variation by age, sex, and socioeconomic position. *Scand J Public Health*. 2011;39(7):669-677. doi:10.1177/1403494811420324

33. Teo YY, Inouye M, Small KS, et al. A genotype calling algorithm for the Illumina BeadArray platform. *Bioinformatics*. 2007;23(20):2741-2746. doi:10.1093/bioinformatics/btm443

34. Delaneau, O.; Marchini, J.; Zagury J. A linear complexity phasing method for thousands of genomes. *Nat Methods*. 2012;(9):179-181. doi:10.1038/nmeth.1785

35. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*. 2009;5(6). doi:10.1371/journal.pgen.1000529

36. Altshuler DL, Durbin RM, Abecasis GR, et al. A map of human genome variation from population-scale sequencing. *Nature*. 2010;467(7319):1061-1073. doi:10.1038/nature09534

37. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet*. 2007;(39):906-913. doi:10.1038/ng2088

38. Palmer TM, Lawlor DA, Harbord RM, et al. Using multiple genetic variants as instrumental variables for modifiable risk factors. *Statistical Methods in Medical Research*. 2012:21:223-242. doi:10.1177/0962280210394459

39. Fewell Z, Davey Smith G, Sterne JAC. The impact of residual and unmeasured confounding in epidemiologic studies: A simulation study. *Am J Epidemiol*. 2007;166(6):646-655. doi:10.1093/aje/kwm165

40. Gupta V, Walia GK, Sachdeva MP. Mendelian randomization: An approach for exploring causal relations in epidemiology. *Public Health*. 2017;145:113-119.

doi:10.1016/j.puhe.2016.12.033

41. von Hinke S, Davey Smith G, Lawlor DA, Propper C, Windmeijer F. Genetic markers as instrumental variables. *J Health Econ*. 2016;45:131-148. doi:10.1016/j.jhealeco.2015.10.007

42. Hemani G, Bowden J, Davey Smith G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum Mol Genet*. 2018;27(R2):R195-R208. doi:10.1093/hmg/ddy163

43. Vanderweele TJ, Tchetgen Tchetgen EJ, Cornelis M, Kraft P. Methodological challenges in Mendelian randomization. *Epidemiology*. 2014;25(3):427-435. doi:10.1097/EDE.0000000000000081

44. Van Kippersluis H, Rietveld CA. Pleiotropy-robust Mendelian randomization. *Int J Epidemiol*. 2018;47(4):1279-1288. doi:10.1093/ije/dyx002

45. Staley JR, Blackshaw J, Kamat MA, et al. PhenoScanner: A database of human genotype-phenotype associations. *Bioinformatics*. 2016;32(20):3207-3209. doi:10.1093/bioinformatics/btw373

46. Lochner L. Nonproduction benefits of education: Crime, health, and good citizenship. In: Hanushek EA, Machin S, Woessmann L, ed. *Handbook of the Economics of Education*. Volume 4. Elsevier; 2011:183-282. doi:10.1016/B978-0-444-53444-6.00002-X

47. Meltzer DO, Jena AB. The economics of intense exercise. *J Health Econ*. 2010. doi:10.1016/j.jhealeco.2010.03.005

48. Sallis JF, Saelens BE. Assessment of physical activity by self-report: Status, limitations, and future directions. *Res Q Exerc Sport*. 2000;71(2):1-14. doi:10.1080/02701367.2000.11082780

49. Davey Smith G. Use of genetic markers and gene-diet interactions for interrogating population-level causal influences of diet on health. *Genes Nutr*. 2011;6(1):27-43. doi:10.1007/s12263-010-0181-y

50. Björklund A, Salvanes KG. Education and family background. Mechanisms and policies. *Handb Econ Educ*. 2011;3:201-247. doi:10.1016/B978-0-444-53429-3.00003-X

51. Brumpton B, Sanderson E, Hartwig FP, et al. Within-family studies for Mendelian randomization: Avoiding dynastic, assortative mating, and population stratification biases. *bioRxiv*. 2019:602516. doi:10.1101/602516

52. Bervoets S, Zenou Y. Intergenerational correlation and social interactions in education. *Eur Econ Rev*. 2017;92(January 2016):13-30. doi:10.1016/j.euroecorev.2016.11.005

53. Becker G. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. Chicago, IL: University of Chicago Press.; 2008.

54. OECD. *Education at a Glance 2018: OECD Indicators*. Paris: OECD Publishing; 2018. doi:10.1787/eag-2018-en

Table 1 Descriptive statistics and comparison of the observables by the instrument value

Panel 1: Descriptive statistics						
	All Mean (SD)[a]	Above median years of education	Below median years of education	Difference	t-statistic	p-value
Overall physical activity in 2011	9.07 (1.89)	9.32 (1.82)	8.94 (1.91)	0.38	3.93	<0.01
Intensive activity, hours/week in 2011	3.50 (1.34)	3.64 (1.25)	3.42 (1.38)	0.22	3.27	<0.01
Total steps per day in 2011	8024 (3042)	7810 (2979)	8141 (3071)	-331	-1.90	0.06
Aerobic steps per day in 2011	1939 (2102)	2158 (2199)	1820 (2039)	338	2.75	<0.01
Education GRS[b]	0.00 (1.00)	0.13 (1.01)	-0.06 (0.98)	0.20	4.55	<0.01
Female (%)	0.56 (0.50)	0.60 (0.49)	0.54 (0.50)	0.06	2.50	0.01
Age (years)	40.81 (5.03)	39.41 (5.07)	41.56 (4.85)	-2.15	-8.45	<0.01
High parental education	0.13 (0.34)	0.24 (0.43)	0.07 (0.26)	0.17	8.76	<0.01
Panel 2: Comparison of the observables by the instrument value						
	All Mean (SD)	Above median GRS	Below median GRS	Difference	t-statistic	p-value
Overall leisure-time physical activity in 2011	9.07 (1.90)	9.19 (1.92)	8.95 (1.85)	0.24	2.60	<0.01
Intensive	3.50	3.54	3.46	0.08	1.18	0.24

activity, hours/week in 2011	(1.34)	(1.32)	(1.35)			
Total steps per day in 2011	8024 (3042)	8174 (3151)	7887 (2933)	287	1.72	0.09
Aerobic Steps per day in 2011	1939 (2102)	1994 (2162)	1889 (2046)	105	0.91	0.37
Education years (2007)	13.88 (2.68)	14.04 (2.74)	13.71 (2.61)	0.32	2.42	0.02
Female (%)	0.56 (0.50)	0.54 (0.50)	0.58 (0.49)	-0.04	-1.73	0.08
Age (years)	40.81 (5.03)	40.69 (5.04)	40.93 (5.03)	-0.24	-0.96	0.34
High parental education	0.13 (0.34)	0.16 (0.37)	0.11 (0.31)	0.05	3.02	<0.01
Other genetic risk scores						
GRS for height	179.90 (8.72)	179.97 (8.53)	179.83 (8.90)	0.14	0.33	0.74
GRS for waist-hip ratio	15.18 (2.36)	15.25 (2.41)	15.12 (2.30)	0.14	1.13	0.26
GRS for BMI	29.10 (3.38)	29.13 (3.35)	29.07 (3.41)	0.06	0.35	0.73

[a] Standard deviation

[b] Genetic Risk Score (unweighted)

Notes: Table reports the means and standard deviations are in parenthesis. Differences between groups were tested using two-sample t test. The indicator for high parental education equals one if at least one of the parents has obtained some university education (based on Longitudinal Population Census data from Statistics Finland).

Table 2 Ordinary Least Squares (OLS) regression results of educational attainment and physical activity

	Overall physical activity in 2011 (n=1651)		Intensive activity, hours / week in 2011 (n=1651)		Total Steps per day in 2011 (n=1338)		Aerobic Steps per day in 2011 (n=1338)	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Per additional year of education	0.07*** (0.02)	0.07*** (0.02)	0.04*** (0.01)	0.04*** (0.01)	-69.60** (31.75)	-80.47** (32.93)	69.69*** (22.34)	59.09*** (22.94)
95% CI	0.04–0.11	0.03–0.10	0.02–0.07	0.02–0.07	-131.89–-7.30	-145.07–-15.87	25.87–113.50	14.08–104.09
R^2	0.03	0.03	0.02	0.02	0.03	0.03	0.07	0.08
Control variables								
Birth cohort, birth month, and gender	x	x	x	x	x	x	x	x
Family Education		x		x		x		x

Notes: Heteroscedasticity-robust standard errors are in parenthesis. Model 1 include controls for gender, cohort (1–6), and birth month. Model 2 include controls for gender, cohort (1–6), birth month, and parents' education. Cohort dummies indicate the year of birth: Cohort 1 = born in 1977, Cohort 2 = born in 1974, Cohort 3 = born in 1971, Cohort 4 = born in 1968, Cohort 5= born in 1965, and Cohort 6 = born in 1962. ***, ** Statistically significant at least at the 1% and 5% levels, respectively.

Table 3 Results of educational attainment and physical activity based on Mendelian Randomization.

	Overall physical activity in 2011 (n=1651)		Intensive activity, hours / week in 2011 (n=1651)		Total Steps per day in 2011 (n=1338)		Aerobic Steps per day in 2011 (n=1388)	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Per additional year of education	0.62*** (0.21)	0.72*** (0.27)	0.26** (0.13)	0.31* (0.16)	559.49* (339.51)	666.11 (423.77)	387.06* (222.49)	428.77 (272.15)
95% CI	0.21–1.03	0.19–1.25	0.01–0.52	-0.002–0.63	-105.93–1224.91	-164.46–1496.68	-49.01–823.13	-104.64–962.17
First-stage F statistic	19.24	13.74	19.24	13.74	13.59	9.92	13.59	9.92
Control variables								
Birth cohort, birth month, and gender	x	x	x	x	x	x	x	x
Family Education		x		x		x		x

Notes: Heteroscedasticity-robust standard errors are in the parentheses. Model 1 include controls for gender, cohort (1–6), and birth month. Model 2 include controls for gender, cohort (1–6), birth month, and family education. Cohort dummies indicate the year of birth: Cohort 1 = born in 1977, Cohort 2 = born in 1974, Cohort 3 = born in 1971, Cohort 4 = born in 1968, Cohort 5= born in 1965, and Cohort 6 = born in 1962. The unweighted GRS is calculated as a sum of genotyped risk alleles or imputed allele dosages carried by an individual, and is standardized with a mean zero and standard deviation for one (28). ***, **, * Statistically significant at least at the 1%, 5% and 10% levels, respectively.

Table 4 Reduced form models

	Overall physical activity in 2011 (n=1651)		Intensive activity, hours / week in 2011 (n=1651)		Total Steps per day in 2011 (n=1338)		Aerobic Steps per day in 2011 (n=1338)	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
GRS (unweighted) for education	0.17*** (0.05)	0.16*** (0.05)	0.07** (0.03)	0.07** (0.03)	142.79* (75.50)	140.96* (75.55)	98.79* (52.99)	90.73* (52.84)
95% CI	0.08–0.26	0.07–0.25	0.01–0.14	0.01–0.13	-5.31–290.90	-7.25–289.17	-5.16–202.73	-12.92–194.39)
R^2	0.02	0.03	0.01	0.01	0.03	0.03	0.07	0.07
Control variables								
Birth cohort, birth month, and gender	x	x	x	x	x	x	x	x
Family Education		x		x		x		x

Notes. Heteroscedasticity-robust standard errors are in the parentheses. Model 1 include controls for gender, cohort (1–6), and birth month. Model 2 include controls for gender, cohort (1–6), birth month, and family education. Cohort dummies indicate the year of birth: Cohort 1 = born in 1977, Cohort 2 = born in 1974, Cohort 3 = born in 1971, Cohort 4 = born in 1968, Cohort 5= born in 1965, and Cohort 6 = born in 1962. The unweighted GRS is calculated as a sum of genotyped risk alleles or imputed allele dosages carried by an individual, and is standardized with a mean zero and standard deviation for one (28). ***, ** Statistically significant at least at the 1% and 5% levels, respectively.

Table 5 Falsification test, adulthood educational attainment and leisure-time self-reported physical activity at 15 years

	Overall physical activity at 15 years		
	OLS	MR (unweighted)	MR (weighted)
Per additional year	0.08***	-0.04	-0.02
of education	(0.02)	(0.20)	(0.19)
95% CI	0.05–0.12	-0.43–0.35	-0.40–0.36
F-value		14.48	15.14
R^2	0.08		
N	1761	1761	1761

Notes. Overall leisure-time physical activity at 15 years is based on same questions than adulthood overall leisure-time physical activity (30–31). Heteroscedasticity-robust standard errors are in parenthesis. All models include controls for gender, cohort (1–6), birth month, and parents' education. Cohort dummies indicate the year of birth: Cohort 1 = born in 1977, Cohort 2 = born in 1974, Cohort 3 = born in 1971, Cohort 4 = born in 1968, Cohort 5= born in 1965, and Cohort 6 = born in 1962. The unweighted GRS is calculated as a sum of genotyped risk alleles or imputed allele dosages carried by an individual, and is standardized with a mean zero and standard deviation for one. The weighted GRS is calculated as a sum of genotyped risk alleles or imputed allele dosages carried by an individual each multiplied by the effect sizes, and it is standardized with a mean zero and standard deviation for one (28).*** Statistically significant at least at the 1% level.