



Educational Attainment Decreases the Risk of COVID-19 Severity in the European Population: A Two-Sample Mendelian Randomization Study

Masahiro Yoshikawa^{1*} and Kensuke Asaba²

¹ Division of Laboratory Medicine, Department of Pathology and Microbiology, Nihon University School of Medicine, Tokyo, Japan, ² Department of Computational Diagnostic Radiology and Preventive Medicine, The University of Tokyo Hospital, Tokyo, Japan

OPEN ACCESS

Edited by:

Bryan Bassig,
National Cancer Institute, National
Institutes of Health (NIH),
United States

Reviewed by:

Alessandra Macciotta,
University of Turin, Italy
Diana A. van der Plaats,
Imperial College London,
United Kingdom
Charles Breeze,
National Institutes of Health (NIH),
United States

*Correspondence:

Masahiro Yoshikawa
myosh-ky@umin.ac.jp

Specialty section:

This article was submitted to
Life-Course Epidemiology and Social
Inequalities in Health,
a section of the journal
Frontiers in Public Health

Received: 01 March 2021

Accepted: 12 April 2021

Published: 03 June 2021

Citation:

Yoshikawa M and Asaba K (2021)
Educational Attainment Decreases the
Risk of COVID-19 Severity in the
European Population: A Two-Sample
Mendelian Randomization Study.
Front. Public Health 9:673451.
doi: 10.3389/fpubh.2021.673451

Observational studies have reported that the severity of COVID-19 depends not only on physical conditions but also on socioeconomic status, including educational level. Because educational attainment (EA), which measures the number of years of schooling, is moderately heritable, we investigated the causal association of EA on the risk of COVID-19 severity using the Mendelian randomization (MR) approach. A two-sample MR analysis was performed using publicly available summary-level data sets of genome-wide association studies (GWASs). A total of 235 single-nucleotide polymorphisms (SNPs) were extracted as instrumental variables for the exposure of EA from the Social Science Genetic Association Consortium GWAS summary data of 766,345 participants of European ancestry. The effect of each SNP on the outcome of COVID-19 severity risk was obtained from the GWAS summary data of 1,059,456 participants of European ancestry gathered from the COVID-19 Host Genetics Initiative. Using inverse variance weighted method, our MR study shows that EA was significantly associated with a lower risk of COVID-19 severity (odds ratio per one standard deviation increase in years of schooling, 0.540; 95% confidence interval, 0.376–0.777, $P = 0.0009$). A series of sensitivity analyses showed little evidence of bias. In conclusion, we show for the first time using a two-sample MR approach the associations between higher EA and the lower risk of COVID-19 severity in the European population. However, the genetic or epidemiological mechanisms underlying the association between EA and the risk of COVID-19 severity remain unknown, and further studies are warranted to validate the MR findings and investigate underlying mechanisms.

Keywords: Mendelian randomization, COVID-19, SARS-CoV-2, educational attainment, years of schooling

INTRODUCTION

The coronavirus disease 2019 (COVID-19), caused by a novel coronavirus SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2), was originally reported as an outbreak of atypical pneumonia cases in Wuhan in the Hubei Province of China in December 2019. As of March 2021, the COVID-19 death toll has topped 2.8 million worldwide according to the World Health Organization (1). Serious COVID-19 patients have pneumonia with hypoxia and may be critical with acute respiratory distress syndrome, pulmonary fibrosis, and other organ failures (2).

Observational studies report that the severity of COVID-19 depends not only on physical conditions such as age, cardiovascular disease, and obesity (3–8) but also on socioeconomic status (SES) indicators such as lower incomes and lower educational level among various populations (9–12). In the European population, lower education level was associated with a higher risk of severe COVID-19 cases that were confirmed either at emergency departments or as inpatients and, therefore, likely reflect severe illness as well as a higher risk of asymptomatic COVID-19 cases in a prospective cohort study using UK Biobank data (9). However, traditional observational studies lacking randomization designs are generally prone to bias by various factors, including confounders and reverse causations (13).

Mendelian randomization (MR) is an epidemiological method that mimics the design of randomized controlled studies using single-nucleotide polymorphisms (SNPs) as instrumental variables (IVs) and examines the causal effects of a risk factor on an outcome of interest. Because genetic variants, such as SNPs, are randomly assigned at conception according to Mendel's law, MR studies are not influenced by confounders or reverse causations and can overcome limitations of observational studies (13). Educational attainment (EA) is highly affected by environmental and social factors but is also moderately heritable as shown by genome-wide association studies (GWASs) (14, 15). Therefore, we were motivated to investigate in this study whether EA had a causal effect on the risk of COVID-19 severity using the MR approach.

METHODS AND MATERIALS

Study Design and Data Sources

We conducted a two-sample MR study using publicly available summary statistics from two GWASs to investigate whether EA was associated with risk of COVID-19 severity. In MR analysis, SNPs from the exposure data set are used as IVs. IVs must satisfy the following three assumptions: The IVs are associated with the exposure (IV assumption 1), the IVs affect the outcome only *via* the exposure (IV assumption 2), and the IVs are not associated with measured or unmeasured confounders (IV assumption 3) (16). For the exposure data set of EA, which measured the number of years of schooling that individuals had completed, the SNPs were obtained from the Social Science Genetic Association Consortium's GWAS summary data of 766,345 participants of European ancestry (13), which was a meta-analysis of 70 discovery cohorts (excluding 23andMe) as shown in **Supplementary Table 1**. This data set was publicly available from the MRC IEU Open GWAS database (17) and MR-Base (18) given as GWAS-ID of "ieu-a-1239." For the outcome data set of the risk of COVID-19 severity, the SNPs were obtained from summary-level GWAS data of COVID-19-hg GWAS meta-analyses (round 5) including 14 studies, but excluding the UK Biobank, with a total of 1,059,456 participants (4,792 very severe respiratory confirmed COVID-19 cases and 1,054,664 controls) of European ancestry by the COVID-19 Host Genetics Initiative (19) (**Supplementary Table 1**), which was released on January 18, 2021, and was also publicly available (20). Very severe respiratory

confirmed COVID-19 cases were defined as hospitalization for laboratory confirmed SARS-CoV-2 infection with death or respiratory support (20).

Selection of Instrumental Variables

The SNPs were selected from the exposure GWAS summary data as IVs by clumping together all SNPs that were associated with EA at a genome-wide significance threshold ($P < 5.0 \times 10^{-8}$) and were not in linkage disequilibrium ($r^2 < 0.01$ and distance $> 10,000$ kb) with the other SNPs. Palindromic SNPs with minor allele frequency > 0.42 were excluded from the analyses (16, 21). As a sensitivity analysis, we also excluded all palindromic SNPs regardless of minor allele frequencies (22). We studied only SNPs that were present in both the exposure and outcome GWAS data sets and did not include proxy SNPs in the analysis (22, 23). To evaluate the strength of the IVs, we calculated the F -statistic of each SNP using the following formula: $F\text{-statistic} = R^2 \times (N-2)/(1-R^2)$, where R^2 is the variance of the phenotype explained by each genetic variant in exposure, and N is the sample size. R^2 was calculated using the following formula: $R^2 = 2 \times (\text{Beta})^2 \times \text{EAF} \times (1-\text{EAF})/[2 \times (\text{Beta})^2 \times \text{EAF} \times (1-\text{EAF}) + 2 \times (\text{SE})^2 \times N \times \text{EAF} \times (1-\text{EAF})]$, where Beta is the per allele effect size of the association between each SNP and phenotype, EAF is the effect allele frequency, and SE is the standard error of Beta (24). IVs with an F -statistic < 10 were regarded as weak instruments (25).

Two-Sample Mendelian Randomization

The Wald ratio, which estimates causal effect for each IV, was calculated as the ratio of Beta for the corresponding SNP in the outcome data set divided by Beta for the same SNP in the exposure data set (26). Our main approach was to conduct a meta-analysis of each Wald ratio by inverse variance weighted (IVW) method using multiplicative random-effects model to estimate overall causal effect of the exposure on the outcome. The causal effects were calculated as the odds ratio (OR) for the risk of COVID-19 severity per one standard deviation (SD) increase in years of schooling (one SD is equivalent to 4.2 years) (15, 27). In addition, we conducted sensitivity analyses by MR-Egger regression, weighted median method, MR-PRESSO (Mendelian Randomization Pleiotropy RESidual Sum and Outlier) global test, and leave-one-out sensitivity analysis. The MR-Egger regression method is used to assess horizontal pleiotropy of IVs. When IV assumption 2 is violated, horizontal pleiotropy occurs, and MR-Egger regression intercept significantly differs from zero (28, 29). The weighted median method provides a valid causal estimate when more than half of the instrumental SNPs satisfy the IV assumptions (24). The MR-PRESSO global test investigates whether there are outlier SNPs whose variant-specific causal estimates differ substantially from those of other SNPs (30, 31). Leave-one-out sensitivity analysis was conducted to assess the reliability of the IVW method by removing each SNP from the analysis and reestimating the causal effect (31). Moreover, among SNPs associated with EA, we searched for SNPs associated with $P < 5.0 \times 10^{-8}$ with pleiotropic effects on body mass index (BMI), smoking, and other SES using the web tool PhenoScanner (version 2) (32, 33). The

heterogeneity was also measured between the causal estimates across all SNPs in the IVW method calculating Cochran’s Q statistic and I^2 statistic (34). Low heterogeneity provides more reliability for a causal effect (35). We conducted all the two-sample MR analyses using “TwoSampleMR” package (version 0.5.5) in *R* (version 4.0.3) (36). A P -value

below 0.05 was considered statistically significant in all statistical analyses.

RESULTS

In total, 235 instrumental SNPs were identified for both EA and the risk of COVID-19 severity GWAS data sets. The characteristics of all the SNPs included in our analysis are shown in **Supplementary Table 2**. The F -statistic of every instrument was >29 , thus suggesting that weak instrument bias was unlikely.

The IVW method showed that EA was significantly associated with a lower risk of COVID-19 severity [OR per 1-SD increase in years of schooling, 0.540; 95% confidence interval (CI), 0.376–0.777; $P = 0.0009$] in the European population (**Table 1**, **Figure 1**, and **Supplementary Figure 1**). Cochran’s Q statistic and I^2 statistic for the IVW method were 261.9 ($P = 0.102$) and 0.110, indicating low heterogeneity and more reliability for the causal effect. Other MR methods also showed overall consistent protective effects for EA on the risk of COVID-19 severity although the MR-Egger regression estimate did not have statistical significance (**Table 1** and **Figure 1**). However, when

TABLE 1 | MR results of the causal effect of EA on the risk of COVID-19 severity.

IVW method	Weighted median method	MR-Egger regression method		Heterogeneity (IVW)	MR-PRESSO global test
OR (95% CI) <i>P</i> -value	OR (95% CI) <i>P</i> -value	OR (95% CI) <i>P</i> -value	Intercept <i>P</i> -value	Cochran’s Q <i>P</i> -value	<i>P</i> -value
0.540 (0.376–0.777) $P = 0.0009$	0.484 (0.283–0.826) $P = 0.008$	0.353 (0.084–1.483) $P = 0.156$	0.006 $P = 0.548$	261.9 $P = 0.102$	$P = 0.115$

MR, Mendelian randomization; COVID-19, coronavirus disease 2019; EA, educational attainment; IVW, inverse variance weighted; OR, odds ratio; CI, confidence interval.

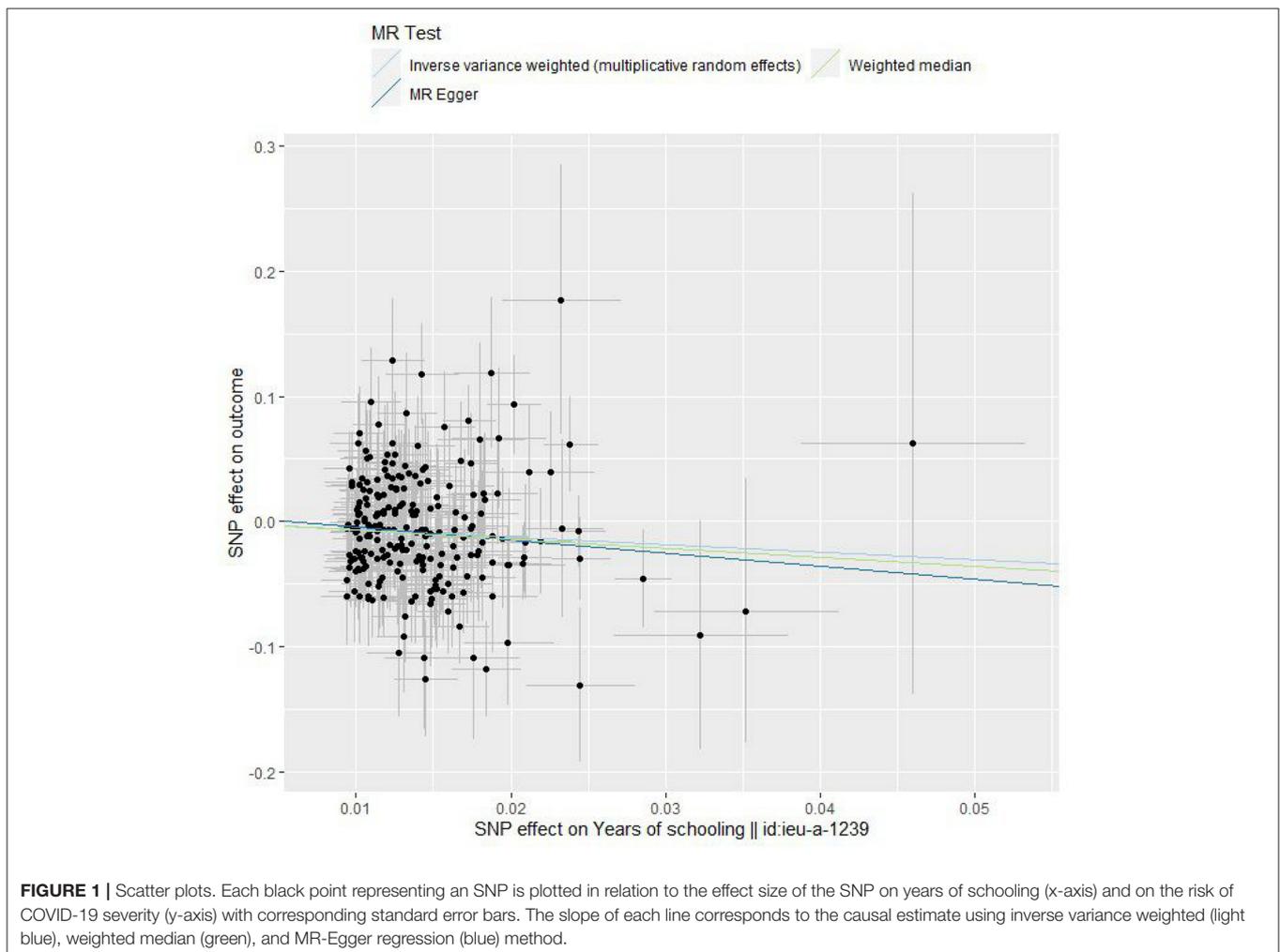


FIGURE 1 | Scatter plots. Each black point representing an SNP is plotted in relation to the effect size of the SNP on years of schooling (x-axis) and on the risk of COVID-19 severity (y-axis) with corresponding standard error bars. The slope of each line corresponds to the causal estimate using inverse variance weighted (light blue), weighted median (green), and MR-Egger regression (blue) method.

I^2 statistics are much <1 , no measurement error assumption is violated and MR-Egger regression tends to underestimate the causal effect (34). In fact, Cochran's Q statistic and I^2 statistic for our MR-Egger regression were 261.5 ($P = 0.097$) and 0.113 (much <1), respectively. The MR-Egger intercept was 0.006 ($P = 0.548$), indicating little evidence of horizontal pleiotropy. The weighted median method indicated that more than half of the instrumental SNPs in our analysis satisfied the IV assumptions. The funnel plot showed general symmetry, suggesting little evidence of heterogeneity or horizontal pleiotropy (**Supplementary Figure 2**). MR-PRESSO global test ($P = 0.115$) and leave-one-out sensitivity analysis suggested the lack of an outlier SNP whose variant-specific causal estimate differed substantially from those of other SNPs (**Table 1** and **Supplementary Figure 3**). We excluded all palindromic SNPs regardless of minor allele frequencies from the IVW method, and then we obtained a comparable result to that of the original IVW method (OR, 0.540; 95% CI, 0.366–0.796, $P = 0.0019$; number of instrumental SNPs; 209). Our search using PhenoScanner identified 17 SNPs that were associated with BMI (rs10073890, rs11123818, rs13090388, rs1334297, rs1566085, rs1618725, rs1689510, rs1964927, rs2725370, rs2820314, rs4787457, rs56391344, rs62444881, rs66568921, rs67890737, rs9372625, rs9384679), three SNPs that were associated with smoking traits (rs10240905 with pack years adult smoking as proportion of life span exposed to smoking, rs2179152 with pack years of smoking, and rs66568921 with ever smoked), and six SNPs that were associated with other SES traits (rs1008078 with Townsend deprivation index at recruitment, rs13090388, rs34316, and rs9372625 with job involving heavy manual or physical work, rs1391438 and rs2971970 with job involving mainly walking or standing), respectively. We excluded the 17 SNPs that were associated with BMI from the IVW method, and then we obtained a comparable result to that of the original IVW method (OR, 0.550; 95% CI, 0.373–0.810, $P = 0.0025$; number of instrumental SNPs; 218) (see section Discussion). When we excluded the three SNPs that were associated with smoking traits from the IVW method, we obtained a comparable result to that of the original IVW method (OR, 0.557; 95% CI, 0.386–0.803, $P = 0.0018$; number of instrumental SNPs; 232). Similarly, we excluded the six SNPs that were associated with other SES traits from the IVW method, and then we obtained a comparable result to that of the original IVW method (OR, 0.536; 95% CI, 0.372–0.773, $P = 0.0009$; number of instrumental SNPs; 229). Moreover, when we excluded all 23 SNPs that were associated with BMI, smoking traits, and other SES traits (three SNPs overlapped), we obtained a comparable result to that of the original IVW method (OR, 0.565; 95% CI, 0.382–0.835, $P = 0.0042$; number of instrumental SNPs; 212).

We noticed some possible overlap between the exposure GWAS participants and the outcome GWAS participants as shown in **Supplementary Table 1**. This might have led to bias in the causal estimate of EA on the risk of COVID-19 severity, but the bias was unlikely to be substantial because the possible overlap was small as discussed below.

DISCUSSION

To our knowledge, this is the first MR study to investigate the association between EA and the risk of COVID-19 severity. Observational studies report that a lower level of education influences the severity of COVID-19 among various populations (9–12). In the European population, those who had no qualification (equivalent to seven years of education) (37) had a higher risk of severe COVID-19 (i.e., a positive test for SARS-CoV-2 in a hospital setting either at emergency departments or as inpatients) than those who had college or university degree (equivalent to 20 years of education) (37) in fully adjusted model [risk ratio (RR), 1.58; 95% CI, 1.25–1.99; $p < 0.001$] in a prospective cohort study using UK Biobank data (9). Our two-sample MR approach supported, with little evidence of bias, the causal effect of higher EA on the risk of COVID-19 severity (OR, 0.540; 95% CI, 0.376–0.777; $P = 0.0009$) in the European population, which was consistent with the cohort study. In other populations, a risk-adjusted model of a large cohort, including 62,298 COVID-19 deaths, showed that lower education levels were strongly associated with the level of COVID-19 fatalities per 100,000 persons (rate ratio, 1.08; 95% CI, 1.05–1.11; $P < 0.0001$) in severely distressed counties in the United States (10). Another study in the United States showed that education level with a bachelor's degree was associated with a lower rate of mortality due to COVID-19 (estimate, -0.246 ; 95% CI, -0.388 to -0.103 ; $P = 0.0008$) across various ethnicities in the seven most affected states (11). In São Paulo, Brazil, among patients under 60 years of age and living in areas with the lowest percentage (below 8.61%) of the population with a university degree, COVID-19 mortality was four times higher than that among those living in areas with the highest percentage (over 34.80%) of population with a university degree (rate ratio, 4.02; 95% CI, 3.42–4.72) (12). However, our MR analysis was based on populations of European ancestry, and the findings are unlikely to be generalized to other populations and ethnicities.

In our MR analysis, underlying genetic or epidemiological mechanisms of how EA lowered the risk of COVID-19 severity remain unknown. Therefore, although a range of sensitivity analyses indicated the robustness of our MR findings, we must pay careful attention to the possibility of unmeasured horizontal pleiotropy of genetic IVs for EA. Observational studies showed that higher EA was associated with decreased prevalence of smoking, physical inactivity, obesity, hypertension, and hypercholesterolemia (38). We infer that other risk factors, including BMI and lifetime smoking, were related to the causal effect of EA on the risk of COVID-19 severity in our analysis for the following reasons. First, MR studies have shown that EA has causal effects on decrease of BMI (39, 40). Second, MR studies have shown that BMI has a causal effect on the risk of COVID-19 severity (29, 41, 42). Consistent with the MR results showing the effect of BMI on the risk of COVID-19 severity, the risk-adjusted model showed that, in addition to the two socioeconomic factors of low level of education and a proportionally larger Black population, obesity was the only physical risk factor in the U.S. cohort (10). Other observational studies also have reported that BMI is a risk factor for hospital

admission, disease severity, and in-hospital mortality due to COVID-19 (5–8). Consistently, our search using PhenoScanner identified 17 SNPs that were associated with both EA and BMI with $P < 5.0 \times 10^{-8}$. Similarly, MR studies have shown that EA has causal effects on increased lifetime smoking (39) and that lifetime smoking has a causal effect on the risk of COVID-19 severity (41) although the associations between smoking and the risk of COVID-19 remain controversial in observational studies (41). Therefore, the causal effect of EA on the risk of COVID-19 severity may be at least partly mediated through increases of BMI and lifetime smoking. Even if that is the case, EA would remain an intervention target for COVID-19 severity (43). In fact, when we excluded the 17 SNPs that were associated with BMI and the three SNPs that were associated with smoking traits, we obtained comparable results to the result of the original IVW method as described above. This supports the idea that EA remains an intervention target for COVID-19 severity because EA lowered the risk of COVID-19 severity to some extent independently of the effects of BMI and smoking.

Epidemiologically, the protective effect of EA on the risk of COVID-19 severity may be related to the social benefit of education. Observational studies showed that lower EA as well as other SES was associated with disparities in medical care (44). For example, counties in the United States with a higher percentage of people below the poverty level had a significantly lower percentage of the population with higher education as well as a lower percentage of people insured (11), and counties in the United States with higher income and education, a lower rate of disability, and a higher rate of the insured population were at a lower risk of COVID-19 mortality (11). However, we must pay attention to interpret the causal association between lower EA and the risk of COVID-19 severity because it remains unclear epidemiologically whether less educated people are more likely to develop severe COVID-19 symptoms. In other words, there is possibility that less educated people are more likely to be socioeconomically disadvantaged and to have an increased risk of SARS-CoV-2 transmission due to poor housing, overcrowding, and low-paid essential jobs that make social distancing more challenging (45). As a result of higher COVID-19 incidence, they may have a higher risk of COVID-19 severity. Ascertainment bias could also arise due to differential healthcare seeking, differential testing, and differential prognosis (9). We could not conduct an MR analysis investigating a causal effect of EA on the risk of COVID-19 incidence as described below. However, the prospective cohort study using UK Biobank data showed that both lower education and area-level socioeconomic deprivation by the Townsend index were associated with having a positive test including asymptomatic COVID-19 [RR 1.46 for no qualifications vs. degree (95% CI 1.19–1.79), and RR 1.39 for most deprived quartile vs. least (95% CI 1.12–1.71)] as well as a higher risk of testing positive in hospital (i.e., severe COVID-19 cases) [RR 1.58 for no qualifications vs. degree (95% CI 1.25–1.99), and RR 1.54 for most deprived quartile vs. least (95% CI 1.21–1.97)] in the fully adjusted model (9). The authors discussed that there remained the possibility that some socioeconomic groups had a poorer prognosis and were, therefore, more likely to be admitted to hospital and, therefore, to be tested (9).

The present study includes the following strengths. First, the samples used were gathered across populations with the same European ancestries, reducing substantial bias in our study. Among different genetic ancestries, effect sizes and allele frequencies can differ and lead to substantial bias (24). Second, we used the publicly available GWAS data sets with the largest sample sizes hitherto for both the exposure and outcome data sets. F -statistics were also large enough for weak instrument bias to be unlikely. Third, a range of sensitivity analyses relaxed the IV assumptions and supported the robustness of our MR findings.

However, we must pay attention to several major limitations. First, in the Geisinger Health System study, the participants in the exposure GWAS may have overlapped with the participants in the outcome GWAS as shown in **Supplementary Table 1**. This might have led to bias in the causal estimate of EA on the risk of COVID-19 severity (46). It was difficult for us to exclude the Geisinger Health System study because we used summary-level data for the exposure and outcome data sets. However, the participants in the Geisinger Health System study represented only 1.9% (14,562 out of 766,344) of those in the exposure GWAS data set. Moreover, the participants in the Geisinger Health System_EUR study represented only 1.2% (53 out of 4,392) of the severe COVID-19 cases, and most of them (10.7%, 112,862 out of 1,054,664) were controls in the outcome GWAS data set. If the data sets are of different sizes, the percentage overlap should be taken with respect to the larger data set (46). Therefore, vast majority of the participant overlap in the outcome GWAS data set occurred, if at all, among the controls. In that situation, the bias is unlikely to be substantial, and unbiased causal estimates are expectedly obtained in two-sample MR studies (41, 46). On the other hand, we could not conduct an MR analysis investigating a causal effect of EA on the risk of COVID-19 incidence because the summary-level GWAS data of COVID-19 incidence (i.e., 32,494 SARS-CoV-2 infection cases and 1,316,207 controls in the European population) by the COVID-19 Host Genetics Initiative (19, 20) had possible participant overlap [at most, 16.2% (5,270 out of 32,494) of the SARS-CoV-2 infection cases in the deCODE_EUR, the Geisinger Health System_EUR, and the Netherlands Twin Register_EUR studies] with the EA GWAS data set that could cause substantial bias (46) (**Supplementary Table 1**). Second, our MR findings might be affected by unmeasured horizontal pleiotropy as described above. As is the often the case with many MR studies, strictly satisfying all the IV assumptions can be challenging (47). Third, our MR analysis was based on populations of European ancestry, and the findings are unlikely to be generalized to other populations and ethnicities. Fourth, we could not conclude that the risk of COVID-19 severity could decrease simply by increasing years of schooling because the underlying genetic or epidemiological mechanisms remain unknown.

In conclusion, we have shown for the first time using a two-sample MR approach the associations between higher EA and the lower risk of COVID-19 severity in the European population that observational studies have reported. However, genetic or epidemiological mechanisms underlying the association between EA and the risk of COVID-19 severity remain unknown, and further studies are warranted to validate our MR findings and investigate underlying mechanisms.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <https://gwas.mrcieu.ac.uk/datasets/ieu-a-1239/>, <https://www.covid19hg.org/results/r5/>.

ETHICS STATEMENT

Ethical approval was not provided for this study on human participants because we used only publicly available GWAS summary datasets. Written informed consent was not provided because we used only publicly available GWAS summary datasets.

AUTHOR CONTRIBUTIONS

MY designed this study, analyzed data, and wrote the draft of the manuscript. MY and KA discussed and reviewed the manuscript critically. Both authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We would like to thank the Social Science Genetic Association Consortium, the MRC IEU Open GWAS database, MR-Base,

and the COVID-19 Host Genetics Initiative for making GWAS datasets publicly available. Our manuscript has been proofread by Editage English Editing Service.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpubh.2021.673451/full#supplementary-material>

Supplementary Figure 1 | Forrest plots. Each black point represents the causal estimate of each SNP on the risk of COVID-19 severity per increase in years of schooling, and red points show the combined causal estimates using IVW and MR-Egger regression methods with horizontal lines denoting 95% confidence intervals.

Supplementary Figure 2 | Funnel plots. Each black point representing an SNP is plotted in relation to the estimate of years of schooling on the risk of COVID-19 severity (x-axis) and the inverse of standard error (y-axis). Vertical lines show the combined causal estimates using IVW (light blue) and MR-Egger regression (blue) methods.

Supplementary Figure 3 | Leave-one-out sensitivity analysis. Each black point represents the combined causal estimates on the risk of COVID-19 severity per increase in years of schooling using IVW methods with horizontal lines denoting 95% confidence intervals after removing the corresponding SNP from the analysis.

Supplementary Table 1 | Contributing studies of the exposure GWAS data and the outcome GWAS data.

Supplementary Table 2 | The characteristics of all the SNPs included in our analysis.

REFERENCES

1. WHO Coronavirus Disease (COVID-19) Dashboard. Available online at: <https://covid19.who.int/> (accessed February 26, 2021).
2. Yuki K, Fujiogi M, Koutsogiannaki S. COVID-19 pathophysiology: a review. *Clin Immunol.* (2020) 215:108427. doi: 10.1016/j.clim.2020.108427
3. Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, et al. Factors associated with COVID-19-related death using Open SAFELY. *Nature.* (2020) 584:430–6. doi: 10.1038/s41586-020-2521-4
4. Koh J, Shah SU, Chua PEY, Gui H, Pang J. Epidemiological and clinical characteristics of cases during the early phase of COVID-19 pandemic: a systematic review and meta-analysis. *Front Med (Lausanne).* (2020) 7:295. doi: 10.3389/fmed.2020.00295
5. Palaodimos L, Kokkinidis DG, Li W, Karamanis D, Ognibene J, Arora S, et al. Severe obesity, increasing age and male sex are independently associated with worse in-hospital outcomes, and higher in-hospital mortality, in a cohort of patients with COVID-19 in the Bronx, New York. *Metabolism.* (2020) 108:154262. doi: 10.1016/j.metabol.2020.154262
6. Popkin BM, Du S, Green WD, Beck MA, Algaith T, Herbst CH, et al. Individuals with obesity and COVID-19: a global perspective on the epidemiology and biological relationships. *Obes Rev.* (2020) 21:e13128. doi: 10.1111/obr.13128
7. Kalligeros M, Shehadeh F, Mylona EK, Benitez G, Beckwith CG, Chan PA, et al. Association of obesity with disease severity among patients with coronavirus disease 2019. *Obesity (Silver Spring).* (2020) 28:1200–4. doi: 10.1002/oby.22859
8. Simonnet A, Chetboun M, Poissy J, Raverdy V, Noulette J, Duhamel A, et al. High prevalence of obesity in severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) requiring invasive mechanical ventilation. *Obesity (Silver Spring).* (2020) 28:1195–9. doi: 10.1002/oby.22831
9. Niedzwiedz CL, O'Donnell CA, Jani BD, Demou E, Ho FK, Celis-Morales C, et al. Ethnic and socioeconomic differences in SARS-CoV-2 infection: prospective cohort study using UK Biobank. *BMC Med.* (2020) 18:160. doi: 10.1186/s12916-020-01640-8
10. Hawkins RB, Charles EJ, Mehaffey JH. Socio-economic status and COVID-19-related cases and fatalities. *Public Health.* (2020) 189:129–34. doi: 10.1016/j.puhe.2020.09.016
11. Abedi V, Olulana O, Avula V, Chaudhary D, Khan A, Shahjouei S, et al. Racial, Economic, and Health Inequality and COVID-19 Infection in the United States. *J Racial Ethn Health Disparities.* (2020) 8:732–42. doi: 10.1101/2020.04.26.20079756
12. Ribeiro KB, Ribeiro AF, de Sousa Mascena Veras MA, de Castro MC. Social inequalities and COVID-19 mortality in the city of São Paulo, Brazil. *Int J Epidemiol.* (2021) dyab022. doi: 10.1093/ije/dyab022. [Epub ahead of print].
13. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet.* (2014) 23:R89–98. doi: 10.1093/hmg/ddu328
14. Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, et al. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature.* (2016) 533:539–42. doi: 10.1038/nature17671
15. Lee JJ, Wedow R, Okbay A, Kong E, Maghziyan O, Zacher M, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet.* (2018) 50:1112–21. doi: 10.1038/s41588-018-0147-3
16. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife.* (2018) 7:e34408. doi: 10.7554/eLife.34408
17. MRC IEU Open GWAS Project. Available online at: <https://gwas.mrcieu.ac.uk/> (accessed February 12, 2021).
18. MR-Base. Available online at: <http://app.mrbase.org/> (accessed February 12, 2021).
19. COVID-19 Host Genetics Initiative. The COVID-19 host genetics initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. *Eur J Hum Genet.* (2020) 28:715–8. doi: 10.1038/s41431-020-0636-6
20. COVID19-hg GWAS Meta-Analyses Round 5. Available online at: <https://www.covid19hg.org/results/r5/> (accessed February 3, 2021).

21. Liu HM, Hu Q, Zhang Q, Su GY, Xiao HM, Li BY, et al. Causal effects of genetically predicted cardiovascular risk factors on chronic kidney disease: a two-sample Mendelian randomization study. *Front Genet.* (2019) 10:415. doi: 10.3389/fgene.2019.00415
22. Gill D, Karhunen V, Malik R, Dichgans M, Sofat N. Cardiometabolic traits mediating the effect of education on osteoarthritis risk: a Mendelian randomization study. *Osteoarthritis Cartilage.* (2021) 29:365–71. doi: 10.1016/j.joca.2020.12.015
23. Xiuyun W, Qian WX, Weidong L, Lizhen L. Education and stroke: evidence from epidemiology and Mendelian randomization study. *Sci Rep.* (2020) 10:21208. doi: 10.1038/s41598-020-78248-8
24. Gill D, Efsthathiadou A, Cawood K, Tzoulaki I, Dehghan A. Education protects against coronary heart disease and stroke independently of cognitive function: evidence from Mendelian randomization. *Int J Epidemiol.* (2019) 48:1468–77. doi: 10.1093/ije/dyz200
25. Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for Mendelian randomization. *Stat Methods Med Res.* (2017) 26:2333–55. doi: 10.1177/0962280215597579
26. Adams CD, Boutwell BB. Can increasing years of schooling reduce type 2 diabetes (T2D)? Evidence from a Mendelian randomization of T2D and 10 of its risk factors. *Sci Rep.* (2020) 10:12908. doi: 10.1038/s41598-020-69114-8
27. Dardani C, Howe LJ, Mukhopadhyay N, Stergiakouli E, Wren Y, Humphries K, et al. Cleft lip/palate and educational attainment: cause, consequence or correlation? A Mendelian randomization study. *Int J Epidemiol.* (2020) 49:1282–93. doi: 10.1093/ije/dyaa047
28. Walker VM, Davies NM, Hemani G, Zheng J, Haycock PC, Gaunt TR, et al. Using the MR-Base platform to investigate risk factors and drug targets for thousands of phenotypes. *Wellcome Open Res.* (2019) 4:113. doi: 10.12688/wellcomeopenres.15334.1
29. Aung N, Khanji MY, Munroe PB, Petersen SE. Causal inference for genetic obesity, cardiometabolic profile and COVID-19 susceptibility: a Mendelian randomization study. *Front Genet.* (2020) 11:586308. doi: 10.3389/fgene.2020.586308
30. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet.* (2018) 50:693–8. doi: 10.1038/s41588-018-0099-7
31. Burgess S, Davey Smith G, Davies NM, Dudbridge F, Gill D, Glymour MM, et al. Guidelines for performing Mendelian randomization investigations. *Wellcome Open Res.* (2020) 4:186. doi: 10.12688/wellcomeopenres.15555.2
32. PhenoScanner V2, A Database of Human Genotype-Phenotype Associations. Available online at: <http://www.phenoscanter.medschl.cam.ac.uk/> (accessed March 25, 2021).
33. Wood A, Guggenheim JA. Refractive error has minimal influence on the risk of age-related macular degeneration: a Mendelian randomization study. *Am J Ophthalmol.* (2019) 206:87–93. doi: 10.1016/j.ajo.2019.03.018
34. Bowden J, Del Greco M F, Minelli C, Davey Smith G, Sheehan NA, Thompson JR. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I² statistic. *Int J Epidemiol.* (2016) 45:1961–74. doi: 10.1093/ije/dyw220
35. Greco MFD, Minelli C, Sheehan NA, Thompson JR. Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. *Stat Med.* (2015) 34:2926–40. doi: 10.1002/sim.6522
36. R Core Team (2020). *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing.
37. Hagenaars SP, Gale CR, Deary IJ, Harris SE. Cognitive ability and physical health: a Mendelian randomization study. *Sci Rep.* (2017) 7:2651. doi: 10.1038/s41598-017-02837-3
38. Hoeymans N, Smit HA, Verkleij H, Kromhout D. Cardiovascular risk factors in relation to educational level in 36 000 men and women in The Netherlands. *Eur Heart J.* (1996) 17:518–25. doi: 10.1093/oxfordjournals.eurheartj.a014903
39. Carter AR, Gill D, Davies NM, Taylor AE, Tillmann T, Vaucher J, et al. Understanding the consequences of education inequality on cardiovascular disease: mendelian randomisation study. *BMJ.* (2019) 365:l1855. doi: 10.1136/bmj.l1855
40. Cao M, Cui B. Association of educational attainment with adiposity, type 2 diabetes, and coronary artery diseases: a Mendelian randomization study. *Front Public Health.* (2020) 8:112. doi: 10.3389/fpubh.2020.00112
41. Li S, Hua X. Modifiable lifestyle factors and severe COVID-19 risk: a Mendelian randomisation study. *BMC Med Genomics.* (2021) 14:38. doi: 10.1186/s12920-021-00887-1
42. Ponsford MJ, Gkatzionis A, Walker VM, Grant AJ, Wootton RE, Moore LSP, et al. Cardiometabolic traits, sepsis, and severe COVID-19: a Mendelian randomization investigation. *Circulation.* (2020) 142:1791–3. doi: 10.1161/CIRCULATIONAHA.120.050753
43. Anderson EL, Howe LD, Wade KH, Ben-Shlomo Y, Hill WD, Deary IJ, et al. Education, intelligence and Alzheimer's disease: evidence from a multivariable two-sample Mendelian randomization study. *Int J Epidemiol.* (2020) 49:1163–72. doi: 10.1093/ije/dyz280
44. Woolf SH, Braveman P. Where health disparities begin: the role of social and economic determinants—and why current policies may make matters worse. *Health Aff (Millwood).* (2011) 30:1852–9. doi: 10.1377/hlthaff.2011.0685
45. Khunti K, Singh AK, Pareek M, Hanif W. Is ethnicity linked to incidence or outcomes of covid-19? *BMJ.* (2020) 369:m1548. doi: 10.1136/bmj.m1548
46. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. *Genet Epidemiol.* (2016) 40:597–608. doi: 10.1002/gepi.21998
47. Burgess S, Butterworth AS, Thompson JR. Beyond Mendelian randomization: how to interpret evidence of shared genetic predictors. *Clin Epidemiol.* (2016) 69:208–16. doi: 10.1016/j.clinepi.2015.08.001

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Yoshikawa and Asaba. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.