Causal Relationship between Educational Attainment and the Risk of Rheumatoid Arthritis: A Mendelian Randomization Study

Guiwu Huang

First Affiliated Hospital of Sun Yat-sen University

Jiahao Cai

Sun Yat-sen Memorial Hospital

Wenchang Li

First Affiliated Hospital of Sun Yat-sen University

Yanlin Zhong

First Affiliated Hospital of Sun Yat-sen University

Weiming Liao

First Affiliated Hospital of Sun Yat-sen University

Peihui Wu (✉ wupeihui3@mail.sysu.edu.cn)

First Affiliated Hospital of Sun Yat-sen University

Research Article

Keywords: Educational attainment, Rheumatoid arthritis, Mendelian randomization Analysis

Posted Date: January 13th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-134634/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background Educational attainment is moderately heritable and positively associated with the risk of rheumatoid arthritis. However, the causality from educational attainment on rheumatoid arthritis remained unknown. Here, we aimed to determine whether educational attainment is causally associated with rheumatoid arthritis (RA) by using a Mendelian randomization (MR) approach.

Methods Summary statistics data for RA were obtained from an available, published meta-analysis of genome-wide association studies (GWAS) that included 14,361 RA cases and 43,923 controls of European ancestry. The instrumental variables for educational attainment were obtained from a GWAS meta-analysis that included over 1 million individuals (N = 1,131,881) of European ancestry. MR analyses were performed using the inverse-variance weighted (IVW), weighted median, and MR-Egger methods. Sensitivity analyses were performed to test the robustness of the association using the Cochran Q test, MR Egger intercept test, “leave-one-out” analysis and MR-PRESSO test.

Results A total of 387 SNPs were employed as instrumental variables in our MR analysis. Genetically predicted higher educational attainment was associated with a significantly lower risk of RA using the IVW method (odds ratio [OR] = 0.42, 95% confidence interval [CI]: 0.34–0.52; $p = 1.78 \times 10^{-14}$). The weighted median and MR Egger methods yielded consistent results. The causality remained robust after discarding the outlier variants and SNPs associated with the confounding factors. “Leave-one-out” analysis confirmed the stability of our results. Additionally, the results demonstrated the absence of the horizontal pleiotropy.

Conclusions The MR analysis supported a potential inverse causative relationship between educational attainment and the risk of RA.

Background

Rheumatoid arthritis (RA) is a systemic autoimmune disease characterized by persistent synovitis, progressive joint disability, and extra-articular manifestations^[1]. It affects up to 1.0% of the population^[2] and is more prevalent and more severe in women^[3]. RA results in poor functional status and chronic pain, which erode the patient’s quality of life, decrease life expectancy and, in some cases, increase mortality^[4, 5], resulting in extra health expenditures of approximately \$19.3 billion per year in the United States^[6, 7].

Although the exact cause of RA remains unclear, both environmental and genetic factors contribute to development of the disease. Previous observational epidemiological studies have shown an association between educational attainment and the risk of RA. However, this association may have arisen from the methodological limitations of traditional observational studies, including residual confounding, reverse causation, and measurement error^[8].

Educational attainment is a well-established socioeconomic and heritable determinant of health^[9], which is measured as number of years of schooling completed (Educational Years). Educational attainment has

been a useful tool in follow-up work to evaluate brain and neural development, biological aging, health behavior and health literacy ^[10, 11]. However, few studies have assessed the association of educational attainment with individual single nucleotide polymorphisms (SNPs).

Mendelian randomization (MR) can be used to evaluate causality by exploiting genetic variants with SNPs as instrumental variables to predict the effect of an exposure on a particular outcome^[12], which overcomes the typical pitfalls such as reverse causation and confounding that hinder observational studies^[13, 14]. Herein, we performed the MR approach to analyze the potential causal effect of educational attainment on the risk of RA.

Methods

Study overview

This study applied MR as a method to determine whether educational attainment is causally associated with RA, using summary data of SNP-exposure (educational attainment) and SNP-outcome (rheumatoid arthritis). An overview of the study design is shown in Figure 1.

GWAS summary statistics

Publicly available summary statistic estimates for the associations between genetic variants and risk of RA were obtained from a meta-analysis of genome-wide association studies (GWAS), including 58,284 individuals from 18 studies of European ancestry (14,361 RA cases and 43,923 controls)^[15]. All RA cases fulfilled the 1987 RA diagnosis criteria of the American College of Rheumatology^[16] or were diagnosed as RA by a professional rheumatologist. Among all RA cases enrolled in this study, 88.1% were seropositive and 9.3% were seronegative for anti-citrullinated peptide antibody (ACPA) or rheumatoid factor (RF), and 2.6% had unknown autoantibody status.

Selection of instrumental variables

The genetic instrumental variables associated with educational attainment were obtained from a GWAS meta-analysis^[17] comprising 71 quality-controlled, cohort-level results files, with a sample size of more than 1 million individuals of European ancestry (N = 1,131,881). All cohort-level analyses were restricted to European-ancestry individuals who passed the cohort's quality control and whose educational years were measured at an age of at least 30. The phenotype was constructed by mapping each major educational qualification that can be identified from the cohort's survey measure to an International Standard Classification of Education (ISCED) category and imputing a years-of-education equivalent for each ISCED category.

A number of quality control steps were taken in our analysis to select eligible instrumental SNPs that were strongly associated with exposure. First, we identified 1,271 lead SNPs at the genome-wide significance threshold ($p < 5 \times 10^{-8}$). After clumping correlated SNPs (linkage disequilibrium [LD] $r^2 > 0.001$), 393 SNPs remained and were used as instrumental variables. We then extracted educational attainment-associated SNPs from the outcome data (RA, in this study). For SNPs absent in the outcome data, we identified proxy SNPs at a cutoff of LD of $r^2 > 0.8$ from the SNIIPA website(<https://snipa.helmholtz-muenchen.de/snipa3/index.php>). SNPs missing in the outcome data without appropriate proxy SNPs available were then excluded. We then calculated the F statistic for each of the SNPs using the following formula:

$$R^2 \times (N - 2)/(1 - R^2)$$

Here, R^2 indicates the proportion of variance in educational attainment explained by a given SNP and N indicates sample size. More specifically, R^2 was calculated with the following formula:

$$R^2 = [2 \times \text{Beta}^2 \times (1 - \text{EAF}) \times \text{EAF}] / [2 \times \text{Beta}^2 \times (1 - \text{EAF}) \times \text{EAF} + 2 \times \text{SE}^2 \times N \times (1 - \text{EAF}) \times \text{EAF}].$$

Here, Beta indicates the genetic effect of SNP on educational attainment, EAF is effect allele frequency, SE is standard error and N is sample size. F statistic is recommended to be over 10 to avoid employing weak genetic instruments^[18].

Statistical analyses

Statistical analyses were performed using the two-sample MR package (version 0.5.5) in R software version 4.0.2 (<https://www.r-project.org/>); $p < 0.05$ was the threshold for a significant difference. All estimates were reported with two-tailed p -values. In the main analysis, we utilized the inverse-variance weighted (IVW), weighted median and MR-Egger methods to investigate the potential causality between educational attainment and RA^[19-21].

To evaluate potential pleiotropy, we further conducted sensitivity analysis using several complementary methods. We firstly evaluated the heterogeneity using Cochran Q test. Then we detected directional pleiotropy using intercept derived from MR-Egger regression^[20]. We also performed leave-one-out analysis to evaluate whether the observed causal relationship was reliant on any single SNP. Finally, MR-PRESSO test was conducted to detect any outlier with potential pleiotropy. Once the outliers were identified, we removed them and repeated MR analysis.

Previous studies have identified smoking initiation^[22], body mass index (BMI)^[23], and type 2 diabetes^[24] as risk factors for RA. Each of the SNPs utilized as an instrumental variable was scanned for its potential

secondary phenotypes using the GWAS catalog (<http://www.ebi.ac.uk/gwas>; accessed on November 17, 2020). Analyses were performed again to test whether the association between educational attainment and RA remained significant after excluding the SNPs associated with traits other than educational attainment.

Results

A total of 1,271 lead SNPs were identified at the genome-wide significance threshold ($p < 5 \times 10^{-8}$) and 393 SNPs remained after clumping ($r^2 < 0.001$) (Additional file, Table 1). However, 7 SNPs were not well matched in the summary statistics data for RA. Of these, one proxy variant (rs11212135, LD $r^2 = 0.82$) was identified for the missing SNP (rs72486027). Six SNPs (rs9859556, rs7029718, rs1334297, rs9375188, rs13018640, rs34305371) without appropriate proxy SNPs available were excluded. Therefore, 387 SNPs were used as instrumental variables for educational attainment in the present study. After the harmonizing process, 14 SNPs were excluded for being non-concordant or palindromic with intermediate allele frequencies (Additional file, Table 2). The SNP rs2256965 was the only outlier variant based on the MR-PRESSO test.

As were shown in **Figure 2**, along with the 373 stringently selected SNPs for subsequent two-sample MR analysis, we found strong evidence to support a causal association between educational attainment and RA using the IVW method (odds ratio [OR] = 0.42, 95% confidence interval [CI]: 0.34–0.52; $p = 1.78 \times 10^{-14}$). The associations were consistent in the sensitivity analyses using the weighted median method (OR = 0.45, 95% CI: 0.34–0.60; $p = 4.27 \times 10^{-8}$). The effect was only slightly attenuated using the MR-Egger method (OR = 0.61; 95% CI: 0.27–1.36; $p = 0.229$). Considering that the weighted median estimator has the advantage of retaining greater precision of the estimates compared with the MR-Egger analysis^[19], the results of the MR analysis support an inversely causative relationship between educational attainment and RA. Furthermore, the association estimated by the IVW method was markedly significant after correction for outlier variants (OR = 0.46, 95% CI: 0.38–0.55, $p = 4.89 \times 10^{-16}$).

Heterogeneity tests suggested an apparent sign of heterogeneity: Q value(df) = 594.21(372), $p = 1.82 \times 10^{-12}$. However, after removing the outlying SNP, heterogeneity was remarkably decreased: Q value(df) = 425.7(371), $p = 0.03$. Additionally, we found no indication of pleiotropy (p -value for Egger intercept = 0.34). The results of the leave-one-out analysis (Additional file, Table 3) demonstrated no potentially influential SNPs driving the causal link between educational attainment and RA in the replication practice.

We then scanned the SNPs for their potential secondary phenotypes using the GWAS catalog. As was shown in Additional file, Table 4, a total of 18 SNPs associated with educational attainment were found to be associated with other traits affecting RA. After excluding SNPs associated with potential confounders, results from the statistical analysis remained essentially consistent (OR = 0.45, 95% CI: 0.37–0.55, $p = 1.05 \times 10^{-15}$, using the IVW method).

Discussion

In this two-sample MR analysis, we found genetic support for a causal association between higher educational attainment and lower risk of RA. On the basis of the intercept estimates of the pleiotropy test, we found no evidence of pleiotropy that likely influenced our results.

In fact, education inequalities in risk of RA have long been noted. Pincus and colleagues^[25] identified an association between a lower level of formal education and higher mortality and morbidity related to RA over a 9-year period. Another study found that formal education level can be a significant marker of clinical status in RA^[26]. However, some studies have revealed that level of formal education is not significantly associated with risk of RA^[27, 28].

Given that these studies with inconsistent conclusions were either based on limited samples or only explored correlations from epidemiological observational studies, few studies have clearly and consistently demonstrated a biological link underlying this association. By applying MR analysis in the current study to alleviate these problems, we provided concrete evidence to support an inverse association of educational attainment with risk of RA. The credibility of this study was verified by using several data sources with large sample sizes. A previous two-sample MR study conducted by Bae and Lee^[29] used statistical data of years of education from the UK Biobank GWAS ($n = 293,723$) as the exposure and a meta-analysis of GWAS of RA with autoantibody ($n = 5,539$) and European controls ($n = 20,169$) as the outcome. Here, we expanded the exposure sample size to over 1 million individuals ($N = 1,131,881$) and chose the latest meta-analysis of GWAS, which included 58,284 individuals of European ancestry (14,361 RA cases and 43,923 controls). Furthermore, individuals with RA who were seropositive and seronegative for ACPA or RF were enrolled in this MR analysis. Thus, the causative association between educational attainment and risk of RA were fully explored in patients with RA.

In total, the identified exposure SNPs accounted for approximately 11% of the variance in educational attainment. The effect size of the independent SNPs corresponding to an educational increase was obtained as follows: the median effect size corresponded to 1.7 weeks of schooling per allele (95% CI: 1.1–2.6 weeks). Furthermore, the genes related to these SNPs are involved in almost all aspects of neuron-to-neuron communication. The dramatic increase in our sample size enabled us to promote the power of the test.

To obtain unbiased estimates, MR needs to fulfill three key assumptions^[12]: (1) genetic variants should be strongly associated with the exposure; (2) genetic variants extracted for exposure should be independent of any confounder; and (3) the genetic variants affect the outcome only through the exposure. The MR approach, which is the closest approximation to a randomized controlled trial, offers one of the most compelling methods to determine causation if there are confounders, because it can minimize the effect of the confounding factors and provide sufficient statistical power for causal estimation.

We used three different methods of estimation for the MR analyses: IVW, weighted median, and MR-Egger method. The IVW and weighted median analyses suggested a negative causal association between educational attainment and RA, whereas the MR-Egger method showed no proof of a causative association between educational attainment and RA. However, the MR-Egger test leads to a loss of precision and power. The weighted median method, which is not influenced by outlying genetic variants, improves the power of causal effect detection and effectively decreases type I error^[19]. Therefore, the weighted median method has a distinct advantage over the MR-Egger test, and its results in this study were the same as those of the IVW method.

The results of our MR analysis might be biased by pleiotropy. Heterogeneity tests suggested an apparent sign of heterogeneity (Q value(df) = 594.21(372), $p = 1.82 \times 10^{-12}$). However, heterogeneity was decreased after removing the outlying SNP (Q value(df) = 425.7(371), $p = 0.03$). Additionally, there was no indication of pleiotropy (p intercept = 0.34). Therefore, we deemed that the conclusion would not be biased significantly by the heterogeneity of the analysis because several robust methods were performed, which can provide reliable inferences and statistical support when some genetic variants violate the assumptions.

The major strength of this study is that the MR design allowed us to investigate the causal association between educational attainment and RA using a large sample size. However, the study has several limitations. The summary GWAS data were restricted to individuals of European descent, and, because ethnicity may affect causality, our results may not be fully representative of the whole population. Another limiting factor was that this applied analysis could not be stratified by sex and age; thus, we could not assess sex discrepancies and potential nonlinear associations.

In conclusion, our aim in this study was to assess the causal effect of educational attainment and risk of RA by using two-sample MR analysis. An updated MR analysis is warranted to confirm our findings of a potential causal association between increased educational attainment and lower risk of RA. These results advocate the current clinical practice for RA surveillance in patients with lower educational attainment.

Abbreviations

MR: Mendelian randomization; GWAS: genome-wide association studies; SNPs: single nucleotide polymorphisms; RA: Rheumatoid arthritis; IVW: Inverse-variance weighted; CI: Confidence interval

Declarations

Ethics approval and consent to participate:

Not applicable

Consent for publication:

Not applicable

Availability of data and materials:

The datasets generated and/or analysed during the current study are publicly available and included in this published article and its supplementary information files.

Competing interests:

Nothing declared.

Funding:

This study was supported by Science and Technology Planning Project of Guangdong Province (2019A141401002), Young Teacher Training Project of Sun Yat-sen University (20ykpy69).

Authors' contributions:

Guiwu Huang and Jiahao Cai collected data, conducted the MR analysis and wrote the manuscript. Peihui Wu and Guiwu Huang contributed to conceptualization, methodology, data acquisition and curation, formal analysis, visualization, writing and editing. Jiahao Cai, Wenchang Li, Yanlin Zhong and Weiming Liao contributed to methodology, interpretation of data, writing and editing. All authors reviewed the manuscript.

Acknowledgments:

The authors acknowledge the participants and investigators of the original studies^[15,17] for sharing the Educational attainment and RA GWAS data used in this study.

Authors' information (optional):

¹ Department of Joint Surgery, The First Affiliated Hospital of Sun Yat-sen University, Sun Yat-sen University, Guangzhou, China.

² Department of Neurology, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, China.

References

1. Scott D, Wolfe F, Huizinga T. Rheumatoid arthritis[J]. *Lancet (London, England)* 2010; 376(9746): 1094–108.
2. Smolen J, Aletaha D, McInnes I. Rheumatoid arthritis[J]. *Lancet (London, England)* 2016; 388(10055): 2023–38.
3. Sokka T, Toloza S, Cutolo M, *et al.* Women, men, and rheumatoid arthritis: analyses of disease activity, disease characteristics, and treatments in the QUEST-RA study[J]. *Arthritis research & therapy* 2009; 11(1): R7.
4. England B, Sayles H, Michaud K, *et al.* Cause-Specific Mortality in Male US Veterans With Rheumatoid Arthritis[J]. *Arthritis care & research* 2016; 68(1): 36–45.
5. Verstappen S. Rheumatoid arthritis and work: The impact of rheumatoid arthritis on absenteeism and presenteeism[J]. *Best practice & research Clinical rheumatology* 2015; 29(3): 495–511.
6. Birnbaum H, Pike C, Kaufman R, *et al.* Societal cost of rheumatoid arthritis patients in the US[J]. *Current medical research and opinion* 2010; 26(1): 77–90.
7. Cross M, Smith E, Hoy D, *et al.* The global burden of rheumatoid arthritis: estimates from the global burden of disease 2010 study[J]. *Annals of the rheumatic diseases* 2014; 73(7): 1316–22.
8. Boyko E. Observational research—opportunities and limitations[J]. *Journal of diabetes and its complications* 2013; 27(6): 642–8.
9. Heath A, Berg K, Eaves L, *et al.* Education policy and the heritability of educational attainment[J]. *Nature* 1985; 314(6013): 734–6.
10. Anttila V, Bulik-Sullivan B, Finucane H, *et al.* Analysis of shared heritability in common disorders of the brain[J]. *Science (New York, NY)* 2018; 360(6395).
11. Marioni R, Ritchie S, Joshi P, *et al.* Genetic variants linked to education predict longevity[J]. *Proceedings of the National Academy of Sciences of the United States of America* 2016; 113(47): 13366–71.
12. Davies N, Holmes M, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians[J]. *BMJ (Clinical research ed)* 2018; 362: k601.
13. Sekula P, Del Greco M F, Pattaro C, *et al.* Mendelian Randomization as an Approach to Assess Causality Using Observational Data[J]. *Journal of the American Society of Nephrology: JASN* 2016; 27(11): 3253–65.
14. Emdin C, Khera A, Kathiresan S. Mendelian Randomization[J]. *JAMA* 2017; 318(19): 1925–6.
15. Okada Y, Wu D, Trynka G, *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery[J]. *Nature* 2014; 506(7488): 376–81.
16. Arnett F, Edworthy S, Bloch D, *et al.* The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis[J]. *Arthritis and rheumatism* 1988; 31(3): 315–24.

17. Lee J, Wedow R, Okbay A, *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals[J]. *Nature genetics* 2018; 50(8): 1112–21.
18. Pierce B, Burgess S. Efficient design for Mendelian randomization studies: subsample and 2-sample instrumental variable estimators[J]. *American journal of epidemiology* 2013; 178(7): 1177–84.
19. Bowden J, Davey Smith G, Haycock P, *et al.* Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator[J]. *Genetic epidemiology* 2016; 40(4): 304–14.
20. Burgess S, Thompson S. Interpreting findings from Mendelian randomization using the MR-Egger method[J]. *European journal of epidemiology* 2017; 32(5): 377–89.
21. Hemani G, Zheng J, Elsworth B, *et al.* The MR-Base platform supports systematic causal inference across the human phenome[J]. *eLife* 2018; 7.
22. Qian Y, Zhang L, Wu D, *et al.* Genetic predisposition to smoking is associated with risk of rheumatoid arthritis: a Mendelian randomization study[J]. *Arthritis research & therapy* 2020; 22(1): 44.
23. Bae S, Lee Y. Causal association between body mass index and risk of rheumatoid arthritis: A Mendelian randomization study[J]. *European journal of clinical investigation* 2019; 49(4): e13076.
24. Inamo J, Kochi Y, Takeuchi T. Is type 2 diabetes mellitus an inverse risk factor for the development of rheumatoid arthritis?[J]. *Journal of human genetics* 2020.
25. Pincus T, Callahan L, Burkhauser R. Most chronic diseases are reported more frequently by individuals with fewer than 12 years of formal education in the age 18–64 United States population[J]. *Journal of chronic diseases* 1987; 40(9): 865–74.
26. Pincus T, Callahan L. Formal education as a marker for increased mortality and morbidity in rheumatoid arthritis[J]. *Journal of chronic diseases* 1985; 38(12): 973–84.
27. Uhlig T, Hagen K, Kvien T. Current tobacco smoking, formal education, and the risk of rheumatoid arthritis[J]. *The Journal of rheumatology* 1999; 26(1): 47–54.
28. Bankhead C, Silman A, Barrett B, *et al.* Incidence of rheumatoid arthritis is not related to indicators of socioeconomic deprivation[J]. *The Journal of rheumatology* 1996; 23(12): 2039–42.
29. Bae S, Lee Y. Causal relationship between years of education and the occurrence of rheumatoid arthritis[J]. *Postgraduate medical journal* 2019; 95(1125): 378–81.

Figures

Figure 1

An overview of study design. IVW, inverse-variance weighted; MR, Mendelian randomization; SNP, single nucleotide polymorphism. *Selection of SNPs: (1) clumping process (threshold: $r^2 = 0.001$, window size = 10,000 kb); (2) excluding SNPs that were not present in the outcome GWAS and cannot be replaced by proxy SNPs; (3) excluding ambiguous and palindromic SNPs, of which the effect cannot be corrected in the harmonizing process; and (4) removing outliers ($p < 0.05$) identified by the MR-PRESSO outlier test.

Figure 2

(A) Scatter plot and (B) Forest plot of Mendelian randomization analyses for the associations of educational attainment with risk of rheumatoid arthritis. OR, odds ratio; CI, confidence interval; IVW, inverse-variance weighted method; MR, Mendelian randomization; SNP, single nucleotide polymorphism.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile.docx](#)