

RESEARCH

Open Access



# Causal relationship between educational attainment and the risk of rheumatoid arthritis: a Mendelian randomization study

Guiwu Huang<sup>1†</sup>, Jiahao Cai<sup>2†</sup>, Wenchang Li<sup>1</sup>, Yanlin Zhong<sup>1</sup>, Weiming Liao<sup>1</sup> and Peihui Wu<sup>1\*</sup>

## Abstract

**Background:** Educational attainment is moderately heritable and inversely 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 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 mainly performed using the inverse-variance weighted (IVW) method. Sensitivity analyses were further performed to test the robustness of the association using the weighted median method, MR-Egger, Cochran Q 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 method and MR Egger regression analysis yielded consistent results. The effect estimate remained robust after the outlier variants and SNPs (associated with the confounding factors) were excluded. “Leave-one-out” analysis confirmed the stability of our results. Additionally, the results suggested the absence of the horizontal pleiotropy.

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

**Keywords:** Educational attainment, Rheumatoid arthritis, Mendelian randomization

## Background

Rheumatoid arthritis (RA) is a systemic autoimmune disease with a lifetime prevalence of 0.5 ~ 1% worldwide [1, 2]. And it's characterized by persistent synovitis, progressive joint disability, and extra-articular manifestations [3]. RA results in poor functional status and chronic pain,

which erodes the patient's quality of life, decreases life expectancy and, in some cases, increases mortality [4, 5], leading to 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 inverse association between educational attainment and the risk of RA. However, this association may be blurred by the methodological limitations of

\* Correspondence: [wupeihui3@mail.sysu.edu.cn](mailto:wupeihui3@mail.sysu.edu.cn)

<sup>†</sup>Guiwu Huang and Jiahao Cai contributed equally to this work.

<sup>1</sup>Department of Joint Surgery, The First Affiliated Hospital of Sun Yat-sen University, Sun Yat-sen University, Guangzhou, China

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

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 defined as years of schooling completed (Edu 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 between educational attainment and RA through genetic variations -- single nucleotide polymorphisms (SNPs).

Mendelian randomization (MR) has been widely 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 confounders 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) based on genome-wide association studies (GWAS). An overview of the study design is shown in Fig. 1.

Assumption 1, genetic variants should be strongly associated with the exposure; Assumption 2, genetic variants extracted for exposure should be independent of any confounder which is associated with both exposure and outcome; and Assumption 3, the genetic variants affect the outcome only through the exposure. \*Sensitivity analyses: Weight median method, MR-Egger regression, MR-PRESSO, Leave-one-out test.

**RA GWAS summary statistics**

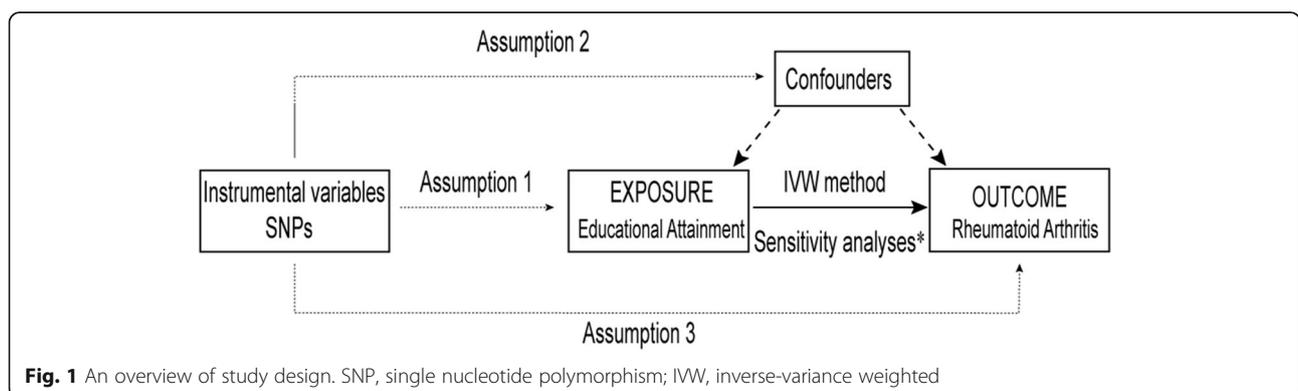
Publicly available summary statistic estimates for the associations between genetic variants and risk of RA were

obtained from a GWAS meta-analysis, 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 (ISCE D) category and imputing a years-of-education equivalent for each ISCED category. In this study, the educational attainment was evaluated by years of schooling completed (Edu Years, mean  $\pm$  standard deviation (SD) =  $16.8 \pm 4.2$  years).

A number of quality control steps were taken in our analysis to select eligible instrumental SNPs that were strongly associated with educational attainment. First, we identified 1271 lead SNPs at the genome-wide significance threshold ( $p < 5 \times 10^{-8}$ ). After clumping correlated SNPs (linkage disequilibrium [LD]  $r^2 \geq 0.001$ ), 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



**Fig. 1** An overview of study design. SNP, single nucleotide polymorphism; IVW, inverse-variance weighted

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) method to investigate the causality between educational attainment and RA [19–21]. And the results were expressed as OR per one SD change in years of education. And one SD in years of education was 4.2 years.

To evaluate potential pleiotropy, weighted median and MR-Egger methods were used as sensitivity analyses. And we detected directional pleiotropy using intercept derived from MR-Egger regression [20]. Then we evaluated the heterogeneity using Cochran Q test. We also performed leave-one-out analysis to evaluate whether the observed causal relationship was reliant on any single SNP. Finally, MR-PRESSO test [22] was conducted to detect any outlier with potential pleiotropy. Once the outliers were identified, we removed them and repeated MR analysis.

Then we retrieved previously published MR studies related to RA from PubMed and identified the potential risk factors causally associated with RA (Vitamin D, body mass index, smoking, alcohol consumption, coffee consumption, mineral nutrients, gut microbiome, diet). These risk factors might be potential confounding factors of this MR study. And confounders considered in this study should be associated with both RA and educational attainment. Therefore, we further conducted a comprehensive search in the GWAS Catalog (<http://www.ebi.ac.uk/gwas>; accessed on November 17, 2020) for whether any SNP in this study was associated with these confounders at the genome-wide significance of  $p < 5 \times 10^{-8}$ . Then we found 18 SNPs associated with confounders (smoking initiation [23], body mass index (BMI) [24], and type 2 diabetes mellitus [25], shown in

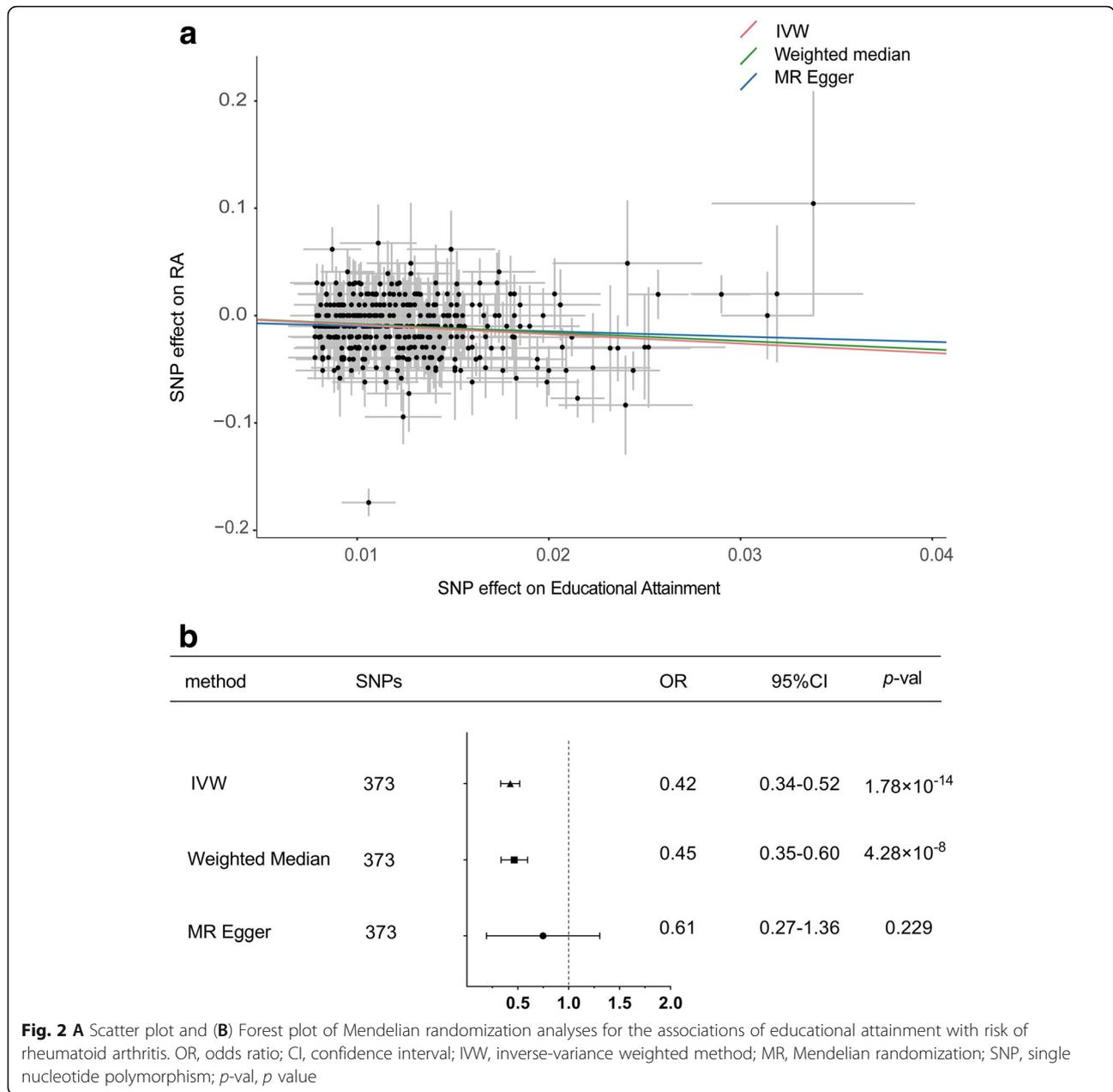
Additional file 1: table S4). As for the other SNPs, we did not find these SNPs were associated with confounding factors. Therefore, we only selected smoking, BMI, and diabetes as confounders in our study. Analyses were performed again to test whether the association remained significant after excluding 18 SNPs associated with RA and risk factors other than educational attainment.

### Results

A total of 1271 lead SNPs were identified at the genome-wide significance threshold ( $p < 5 \times 10^{-8}$ ) and 393 SNPs remained after clumping ( $r^2 < 0.001$ ). However, seven SNPs were not available 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 (rs11657342, rs182902112, rs73581580, rs75033012, rs75177132, rs76246107) without appropriate proxy SNPs available were excluded. And after the harmonizing process, one SNP (rs77719387) was removed for incompatible alleles and 13 SNPs were excluded for being palindromic with intermediate allele frequencies (Additional file, Table 1). Therefore, 373 SNPs were chosen as instrumental variables for educational attainment in the present study (Additional file, Table 2). The SNP rs2256965 was the only outlier variant detected by the MR-PRESSO test.

As were shown in Fig. 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;  $p = 1.78 \times 10^{-14}$ ), which meant that RA risk decreased by 58% per SD (approximately 4.2 years) increased in the years of schooling. The associations were consistent in the sensitivity analysis 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 had the advantage of retaining greater precision of the estimates compared with the MR-Egger analysis [19] and MR-Egger method was often used as a reference for the direction of causal association the results of the MR analysis, we believe that our study supports an inversely causative relationship between educational attainment and RA. Furthermore, the association estimated by the IVW method was markedly significant after correction for one outlier variant detected by MR-PRESSO test (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),  $p =$



**Fig. 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; *p*-val, *p* value

0.03. Additionally, we found no indication of unbalanced 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 replicated analyses.

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 leveraged the largest genetic data set for educational attainment published to date, together with the largest GWAS addressing the outcome of interest, to understand causal relationships between educational attainment and risk of RA. We identified a pronounced causal effect that genetic predisposition to higher educational attainment was associated with a lower risk of RA.

In fact, education inequalities in risk of RA have long been noted. Pincus and colleagues [26] 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 [27]. However, some studies have revealed that level of formal education is not significantly associated with risk of RA [28, 29].

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 causal association of educational attainment with risk of RA. The credibility of this study was verified by using several data sets with largest sample size.

A previous two-sample MR study conducted by Bae and Lee [30] suggested that RA risk decreased by 52% per SD (approximately 3.61 years) increased in the years of schooling completed used relatively small sample size statistical data set of years of education from the UK Biobank GWAS ( $n = 293,723$ ) as the exposure and a meta-analysis of GWAS of RA ( $n = 5539$ ) 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). Meanwhile, the number of SNPs chosen as instrumental variables increased dramatically (from 49 to 373). Furthermore, individuals with RA who were seropositive and seronegative for ACPA or RF were enrolled in this MR analysis. Recently, a comparable analysis has been carried out by Yuan et al. used the same GWAS as sources for educational attainment (EA, exposure) and rheumatic arthritis (RA, outcome) [31]. And they found that RA risk decreased by 50% per SD (approximately 4.2 years) increased in the years of schooling completed. However, there are some differences that should be noticeable between our study and the Yuan's. Firstly, the linkage disequilibrium (LD)  $r^2$  of the SNPs in our study was set to a more conservative threshold to obtain a smaller set of SNPs (LD  $r^2 < 0.001$  with 387 SNPs remained vs. LD  $r^2 < 0.01$  with 663 SNPs used in Yuan's study) to ensure the independence of SNPs at the cost of decreasing statistical power. Secondly, the RA phenotype used in Yuan's study came from mixed populations (European and Asian ancestry). However, it should be noticed that population heterogeneity may lead to bias to the MR results. As such, all the association analyses performed in our study were restricted to European-

descent individuals, making the MR estimates more reliable. As such, our MR study is based on merely European descents and a set of more conservative genetic instruments, to appraise the causal relationship between EA and the risk of RA. Thus, the causative association fully explored in patients between educational attainment and risk of RA was more convincing.

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 [17]. The dramatic increase in our sample size enabled us to improve the power of the test.

The MR approach, as an approximation to a randomized controlled trial in nature, offers one of the most compelling methods to detect causation. The IVW method and weighted median method suggested an inverse 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 provided a reference for the direction of causal association. The weighted median method, which is not influenced by outlying genetic variants, improved the power of causal effect detection and effectively decreased type I error [19]. Therefore, the weighted median method had a distinct advantage over the MR-Egger test, and its result in this study was the same as that 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 detected by MR PRESSO test (Q value (df) = 425.7(371),  $p = 0.03$ ). Additionally, there was no indication of unbalanced pleiotropy ( $p$ -value for MR Egger 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 could provide reliable inferences and statistical support.

The potential mechanisms that educational attainment ultimately reduces the risk of RA may be complicated. In general, higher educational attainment is associated with greater wealth and status, as well as healthier lifestyle and relatively higher quality of life. This cascade of benefits from higher educational attainment may eventually contribute to RA prevention. In addition, the effect of educational attainment on RA may also be mediated by obesity. In an MR study, Böckerman et al. found that

the higher years of schooling was associated with lower BMI [32]. Another MR study has also shown that higher years of schooling was associated with lower plasma triglyceride levels, waist circumference and waist-to-hip ratio [33]. This study suggests that higher education attainment can protect against obesity to some extent and higher educational attainment could be a protective factor against obesity in advanced countries [32]. What's more, previous study has shown that lower educational attainment was associated with an increased risk of Type 2 Diabetes Mellitus (T2DM) [33]. And T2DM was a significant risk factor for RA [25, 34]. And Qian et al. has found that genetic predisposition to smoking was positively associated with rheumatoid arthritis [23]. To some extent, some studies have found that well-educated individuals were less likely to smoke [35–37]. Based on the points discussed above, the influence of educational attainment on individuals has multiple aspects, and the total vector effects reflected in RA is the preventive effect. However, the potential biological mechanisms were rarely reported. Further researches are needed to uncover the pathways about how the educational attainment decreases the risk of RA.

This 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 non-European populations. Another limiting factor was that this applied analysis could not be stratified by gender and age due to the meta-GWAS was performed without adjustment for gender or age to maximize statistical power, thus, we could not assess gender or age 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 with pretty large sample sizes. However, further confirmatory methods should be conducted to verify 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 those with lower educational attainment.

#### Abbreviations

MR: Mendelian randomization; GWAS: Genome-wide association studies; SNPs: Single nucleotide polymorphisms; RA: Rheumatoid arthritis; IWW: Inverse-variance weighted; CI: Confidence interval

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s41927-021-00216-0>.

**Additional file 1: Table S1.** Detailed information of SNPs Harmonizing. **Table 2.** Detailed information of LD-independent SNPs chosen as instrumental variables for educational attainment (exp) and rheumatoid arthritis (out). **Table 3.** Detailed information of the “leave-one-out” analysis

corresponding to the IWW analysis. **Table 4.** The SNPs and their corresponding phenotypes considered as confounders.

#### 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' contributions

HGW and CJH collected data, conducted the MR analysis and wrote the manuscript. WPH and HGW contributed to conceptualization, methodology, data acquisition and curation, formal analysis, visualization, writing and editing. CJH, LWC, ZYL and LWM contributed to methodology, interpretation of data, writing and editing. All authors reviewed the manuscript. The authors read and approved the final manuscript.

#### Funding

This study was supported by Science and Technology Planning Project of Guangdong Province (2017B020227005). Science and Technology Planning Project of Guangdong Province (2019A141401002), Young Teacher Training Project of Sun Yat-sen University (20ykpy69).

#### 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.

#### Declarations

##### Ethics approval and consent to participate

Not applicable.

##### Consent for publication

Not applicable.

##### Competing interests

Nothing declared.

##### Author details

<sup>1</sup>Department of Joint Surgery, The First Affiliated Hospital of Sun Yat-sen University, Sun Yat-sen University, Guangzhou, China. <sup>2</sup>Department of Neurology, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, China.

Received: 23 December 2020 Accepted: 3 July 2021

Published online: 21 October 2021

#### References

- Aletaha D, Smolen JS. Diagnosis and Management of Rheumatoid Arthritis: a review. *JAMA*. 2018;320(13):1360–72. <https://doi.org/10.1001/jama.2018.13103>.
- Smolen JS, Aletaha D, McInnes IB. Rheumatoid arthritis. *Lancet*. 2016; 388(10055):2023–38. [https://doi.org/10.1016/S0140-6736\(16\)30173-8](https://doi.org/10.1016/S0140-6736(16)30173-8).
- Scott D, Wolfe F, Huizinga T. Rheumatoid arthritis. *Lancet*. 2010;376(9746): 1094–108.
- England B, Sayles H, Michaud K, et al. Cause-specific mortality in male US veterans with rheumatoid arthritis. *Arthritis Care Res*. 2016;68(1):36–45. <https://doi.org/10.1002/acr.22642>.
- Verstappen S. Rheumatoid arthritis and work: the impact of rheumatoid arthritis on absenteeism and presenteeism. *Best Pract Res Clin Rheumatol*. 2015;29(3):495–511. <https://doi.org/10.1016/j.berh.2015.06.001>.
- Birnbaum H, Pike C, Kaufman R, Maynchenko M, Kidolezi Y, Cifaldi M. Societal cost of rheumatoid arthritis patients in the US. *Curr Med Res Opin*. 2010;26(1):77–90. <https://doi.org/10.1185/03007990903422307>.
- Cross M, Smith E, Hoy D, Carmona L, Wolfe F, Vos T, et al. The global burden of rheumatoid arthritis: estimates from the global burden of disease 2010 study. *Ann Rheum Dis*. 2014;73(7):1316–22. <https://doi.org/10.1136/annrheumdis-2013-204627>.
- Boyko E. Observational research—opportunities and limitations. *J Diabetes Complicat*. 2013;27(6):642–8. <https://doi.org/10.1016/j.jdiacomp.2013.07.007>.

9. Heath A, Berg K, Eaves L, et al. Education policy and the heritability of educational attainment. *Nature*. 1985;314(6013):734–6. <https://doi.org/10.1038/314734a0>.
10. Anttila V, Bulik-Sullivan B, Finucane H, et al. Analysis of shared heritability in common disorders of the brain. *Science*. 2018;360:6395.
11. Marioni R, Ritchie S, Joshi P, et al. Genetic variants linked to education predict longevity. *Proc Natl Acad Sci U S A*. 2016;113(47):13366–71. <https://doi.org/10.1073/pnas.1605334113>.
12. Davies N, Holmes M, Davey SG. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ*. 2018;362:k601.
13. Sekula P, Del Greco MF, Pattaro C, et al. Mendelian randomization as an approach to assess causality using observational data. *J Am Soc Nephrol*. 2016;27(11):3253–65. <https://doi.org/10.1681/ASN.2016010098>.
14. Emdin C, Khera A, Kathiresan S. Mendelian randomization. *JAMA*. 2017; 318(19):1925–6. <https://doi.org/10.1001/jama.2017.17219>.
15. Okada Y, Wu D, Trynka G, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature*. 2014;506(7488):376–81. <https://doi.org/10.1038/nature12873>.
16. Arnett F, Edworthy S, Bloch D, et al. The American rheumatism association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum*. 1988;31(3):315–24. <https://doi.org/10.1002/art.1780310302>.
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. *Nat Genet*. 2018;50(8):1112–21. <https://doi.org/10.1038/s41588-018-0147-3>.
18. Pierce B, Burgess S. Efficient design for Mendelian randomization studies: subsample and 2-sample instrumental variable estimators. *Am J Epidemiol*. 2013;178(7):1177–84. <https://doi.org/10.1093/aje/kwt084>.
19. Bowden J, Davey Smith G, Haycock P, et al. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol*. 2016;40(4):304–14. <https://doi.org/10.1002/gepi.21965>.
20. Burgess S, Thompson S. Interpreting findings from Mendelian randomization using the MR-egger method. *Eur J Epidemiol*. 2017;32(5): 377–89. <https://doi.org/10.1007/s10654-017-0255-x>.
21. 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. <https://doi.org/10.7554/eLife.34408>.
22. 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(5):693–8. <https://doi.org/10.1038/s41588-018-0099-7>.
23. Qian Y, Zhang L, Wu D, et al. Genetic predisposition to smoking is associated with risk of rheumatoid arthritis: a Mendelian randomization study. *Arthritis Res Ther*. 2020;22(1):44. <https://doi.org/10.1186/s13075-020-2134-1>.
24. Bae S, Lee Y. Causal association between body mass index and risk of rheumatoid arthritis: a Mendelian randomization study. *Eur J Clin Invest*. 2019;49(4):e13076. <https://doi.org/10.1111/eci.13076>.
25. Inamo J, Kochi Y, Takeuchi T. Is type 2 diabetes mellitus an inverse risk factor for the development of rheumatoid arthritis? *J Hum Genet*. 2020;66:219–23.
26. 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 Chronic Dis*. 1987;40(9):865–74. [https://doi.org/10.1016/0021-9681\(87\)90186-X](https://doi.org/10.1016/0021-9681(87)90186-X).
27. Pincus T, Callahan L. Formal education as a marker for increased mortality and morbidity in rheumatoid arthritis. *J Chronic Dis*. 1985;38(12):973–84. [https://doi.org/10.1016/0021-9681\(85\)90095-5](https://doi.org/10.1016/0021-9681(85)90095-5).
28. Uhlig T, Hagen K, Kvien T. Current tobacco smoking, formal education, and the risk of rheumatoid arthritis [J]. *J Rheumatol*. 1999;26(1):47–54.
29. Bankhead C, Silman A, Barrett B, Scott D, Symmons D. Incidence of rheumatoid arthritis is not related to indicators of socioeconomic deprivation. *J Rheumatol*. 1996;23(12):2039–42.
30. Bae S, Lee Y. Causal relationship between years of education and the occurrence of rheumatoid arthritis. *Postgrad Med J*. 2019;95(1125):378–81. <https://doi.org/10.1136/postgradmedj-2018-136374>.
31. Yuan S, Xiong Y, Michaëlsson M, Michaëlsson K, Larsson SC. Genetically predicted education attainment in relation to somatic and mental health. *Sci Rep*. 2021;11(1):4296. <https://doi.org/10.1038/s41598-021-83801-0>.
32. Böckerman P, Viinikainen J, Pulkki-Råback L, Hakulinen C, Pitkänen N, Lehtimäki T, et al. Does higher education protect against obesity? Evidence using Mendelian randomization. *Prev Med*. 2017;101:195–8. <https://doi.org/10.1016/j.ypmed.2017.06.015>.
33. Liao LZ, Chen ZC, Li WD, Zhuang XD, Liao XX. Causal effect of education on type 2 diabetes: a network Mendelian randomization study. *World J Diabetes*. 2021;12(3):261–77. <https://doi.org/10.4239/wjcd.v12i3.261>.
34. Jiang P, Li H, Li X. Diabetes mellitus risk factors in rheumatoid arthritis: a systematic review and meta-analysis. *Clin Exp Rheumatol*. 2015;33(1):115–21.
35. Kim JH, Noh J, Choi JW, et al. Association of education and smoking status on risk of diabetes mellitus: a population-based nationwide cross-sectional study. *Int J Environ Res Public Health*. 2017;14(6):655.
36. de Walque D. Does education affect smoking behaviors? Evidence using the Vietnam draft as an instrument for college education. *J Health Econ*. 2007;26(5):877–95. <https://doi.org/10.1016/j.jhealeco.2006.12.005>.
37. Lawrence EM. Why do College graduates behave more healthfully than those who are less educated? *J Health Soc Behav*. 2017;58(3):291–306. <https://doi.org/10.1177/0022146517715671>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

