OPEN ACCESS

Examining causal relationships between educational attainment and type 2 diabetes using genetic analysis: findings from the EPIC-InterAct study through Mendelian randomisation

Alessandra Macciotta ,[1,2] Carlotta Sacerdote,[3] Claudia Giachino,[1] Chiara Di Girolamo,[1] Matteo Franco,[1] Yvonne T van der Schouw,[4] Raul Zamora-Ros,[5] Elisabete Weiderpass,[6] Cloé Domenighetti,[7] Alexis Elbaz ,[7] Thérèse Truong,[7] Claudia Agnoli,[8] Benedetta Bendinelli,[9] Salvatore Panico,[10] Paolo Vineis ,[11] Sofia Christakoudi,[11,12] Matthias B Schulze,[13,14,15] Verena Katzke,[16] Rashmita Bajracharya,[16] Christina C Dahm,[17] Susanne Oksbjerg Dalton,[18,19] Sandra M Colorado-Yohar,[20,21,22] Conchi Moreno-Iribas,[23] Pilar Amiano Etxezarreta,[21,24,25] María José Sanchez,[21,26,27] Nita G Forouhi,[28] Nicholas Wareham,[28] Fulvio Ricceri [1]

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/jech-2024-222734).

For numbered affiliations see end of article.

Correspondence to
Professor Carlotta Sacerdote; carlotta.sacerdote@uniupo.it

NW and FR contributed equally.

Received 10 July 2024
Accepted 19 November 2024
Published Online First 10 December 2024

To cite: Macciotta A, Sacerdote C, Giachino C, *et al*. *J Epidemiol Community Health* 2025;**79**:373–379.

ABSTRACT

Introduction Observational studies have shown that more educated people are at lower risk of developing type 2 diabetes (T2D). However, robust study designs are needed to investigate the likelihood that such a relationship is causal. This study used genetic instruments for education to estimate the effect of education on T2D using the Mendelian randomisation (MR) approach.

Methods Analyses have been conducted in the European Prospective Investigation into Cancer and Nutrition (EPIC)-InterAct study (more than 20 000 individuals), a case-cohort study of T2D nested in the EPIC cohort. Education was measured as Years of Education and Relative Index of Inequality. Prentice-weighted Cox models were performed to estimate the association between education and T2D. One-sample MR analyses investigated whether genetic predisposition towards longer education was associated with risk of T2D and investigated potential mediators of the association.

Results MR estimates indicated a risk reduction of about 15% for each year of longer education on the risk of developing T2D, confirming the protective role estimated by observational models (HR 0.96, 95% CI 0.95 to 0.96). MR analyses on putative mediators showed a significant role of education on body mass index, alcohol consumption, adherence to the Mediterranean diet and smoking habits.

Conclusion The results supported the hypothesis that higher education is a protective factor for the risk of developing T2D. Based on its position in the causal chain, education may be antecedent of other known risk factors for T2D including unhealthy behaviours. These findings reinforce evidence obtained through observational study designs and bridge the gap between correlation and causation.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Observational studies showed an inverse association between educational attainment and type 2 diabetes (T2D).
⇒ Mendelian randomisation (MR) can overcome typical bias of observational studies and can be used to infer causality.
⇒ Two-sample MR studies increased evidence of the causality of this association.

WHAT THIS STUDY ADDS

⇒ One-sample MR showed an inverse association between educational attainment and T2D incidence.
⇒ The results are consistent using different indices to measure educational attainment.
⇒ Potential mediators of this association should be sought in body mass index and in smoking, drinking and eating habits.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ The findings may drive preventive policies: targeting both people more at risk and potential mechanisms able to explain the inequalities, the burden of T2D may be reduced.

INTRODUCTION

Type 2 diabetes (T2D) is one of the most concerning health issues worldwide. Recent studies[1] estimated that more than 6% of the world's population is currently affected by T2D and both its incidence and prevalence are expected to increase in the future. T2D often results in severe complications, requiring expensive and long-lasting treatments.[2] Since T2D is a chronic condition, preventive measures play a key role in reducing its burden on public health. Such preventive interventions need to be based on a sound understanding of the determinants of T2D.

Observational studies[3 4] have identified socioeconomic position (SEP) as an important determinant

of T2D. These studies show that in high-income countries individuals with lower SEP are more likely to develop T2D than those with higher SEP. Possible explanations for this association may link to differential exposure, detrimental behaviours and stressors, as well as differential access to healthcare services.[5] In observational studies, however, it is often difficult to discriminate between association and causation due to the potential for such studies to be affected by confounding and reverse causality.

Randomised clinical trials (RCTs) have traditionally been considered the gold standard approach for the investigation of causality since random allocation of individuals to the exposure of interest reduces the likelihood of both measured and unmeasured confounding.[6] However, RCTs are rarely feasible in the context of social epidemiology. The Mendelian randomisation (MR) approach takes advantage of the random allocation at meiosis of genetic variants which can be used as a proxy for the exposure and can facilitate the investigation of potential causal relationships in observational studies.[7]

Two Genome-Wide Association Studies (GWAS)[8 9] performed in the last decade identified several single-nucleotide polymorphisms (SNPs) associated with educational attainment, which is one of the most stable proxies for SEP.[10]

Starting from summary statistics obtained by these two GWAS,[8 9] several researchers have explored the causal relationship between educational attainment and T2D[11–16] mainly through two-sample MR designs. Results yielded different conclusions: some found a protective effect of higher education on T2D,[11–14] while others reported null associations.[15 16]

In this study, to test and estimate the causal effect of educational attainment on T2D incidence, we combined more than 1000 genetic variants to predict educational attainment and then we performed one-sample MR analyses. To the best of our current knowledge, our analyses are the first use of MR using individual-level data to investigate this association considering the genetic variants discovered by Lee *et al*.[9] Indeed, one-sample MR analyses were performed only in a recently published paper,[17] where only genetic variants obtained by the less recent GWAS on educational attainment[8] were used. We also tested MR models considering putative mediators of the association between educational attainment and T2D including smoking and drinking habits, adherence to the Mediterranean diet, body mass index (BMI) and physical activity.

METHODS

Study participants

The analyses were performed using data from the European Prospective Investigation into Cancer and Nutrition (EPIC)-InterAct study, a case-cohort nested in the EPIC cohort.[18] The EPIC cohort consists of more than 500 000 volunteers recruited between 1992 and 2000 in 10 European countries. Baseline information regarding lifestyle behaviours, indicators for SEP and pre-existing clinical conditions were collected using standardised questionnaires. Blood samples were collected and stored at recruitment for about 380 000 participants.

The EPIC-InterAct study was designed to investigate how genetic and behavioural risk factors affect the risk of developing T2D.[19] 12 403 cases of incident T2D from 8 of the 10 EPIC countries were included (Denmark, France, Germany, Italy, Spain, Sweden, The Netherlands and UK), together with a randomly selected subcohort consisting of about 16 000 individuals. As a feature of a case-cohort study, there was an overlap of 778 cases in the cases and the subcohort.

Selection of SNPs and definition of genetic-risk scores

Genotyping in EPIC-InterAct was performed using the Illumina Human Core Exome array and the Illumina Human Quad 660 array.

1259 out of the 1271 SNPs identified by Lee *et al*[9] were available in the EPIC-InterAct case cohort. 102 palindromic SNPs with an effect allele frequency between 0.3 and 0.7 were excluded, and thus 1157 SNPs (online supplemental table 1) were included to build a genetic risk scores (GRS), namely a linear combination of risk alleles and external weights. Formally, each participant i with g copies of the risk alleles k with a specific weight w_k was attributed a GRS equal to $\sum_{k=0}^{K} w_k g_{ik}$.

Since using weights estimated from independent cohorts which are similar and well powered is thought to be the best approach for building weighted allele scores,[20] we considered as weights the effect estimates computed in the original GWAS.[9]

Definition of SEP

SEP was measured as educational attainment as the number of years of schooling completed, a similar phenotype to that studied by the original GWAS.[9] In detail, we transformed the available variable indicating the highest school level achieved using the International Standard Classification of Education (ISCED) and imputed a new variable, years of education (YoE), equivalent to each ISCED value.

Since the EPIC-InterAct case-cohort consists of men and women born in different European countries throughout the twentieth century, we computed Relative Index of Inequality (RII), which allows educational levels to be compared.[21] Each individual was assigned a value between 0 and 1, depending on the proportion of individuals with their own educational level in the subgroup of participants from the same country of origin, sex, and age (10-year intervals). We then computed (1-RII) times 20, referred to as transformed RII (RII_t), to make it comparable with the YoE variable. Indeed 20 is the maximum value for YoE.

Definition of putative mediators

We investigated which risk factors were identified in the literature as putative mediators of the association between education and T2D and selected those available in the EPIC-InterAct database. We considered continuous measurements for BMI, daily alcohol consumption (grams), Mediterranean diet score[22] and dichotomous measurements for smoking habits (never vs former/current) and physical activity[23] (inactive/moderately inactive vs moderately active/active).

Descriptive analysis

We used frequencies and percentages to describe categorical variables, while we used means and SDs for continuous variables.

Observational analysis

The separate observational associations between YoE or RII_t and T2D were estimated by Prentice-weighted Cox regressions in the case cohort, using days of follow-up as the underlying time scale. We adjusted the first model considering YoE as exposure with age, sex and country, while none covariate was added for the model with RII_t as the exposure since these have been accounted for when calculating the index.

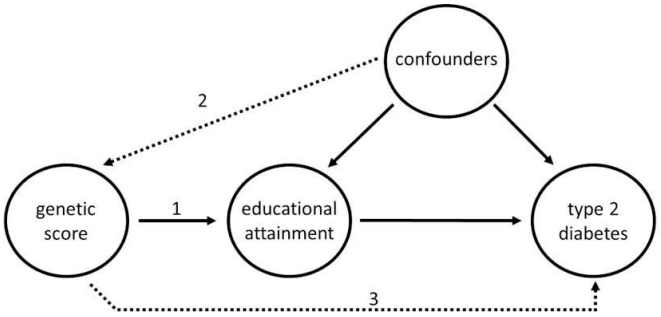

Figure 1 Directed acyclic graph depicting the hypothesised connections among variables in the Mendelian randomisation (MR) analysis variables in the EPIC-InterAct study. Solid lines represent the associations to be fulfilled, while dotted lines represent the associations which are supposed to be absent to perform MR. The numbers denote the MR assumptions as described in the main text. EPIC, European Prospective Investigation into Cancer and Nutrition.

Furthermore, we estimated observational association between YoE or RII_t and possible mediators. We performed linear and logistic regressions, according to the distribution of each single mediator, on the EPIC-Interact subcohort.

MR analyses

The MR consists of using a genetic variable, GRS in our study, as an instrumental variable to mimic a randomisation process for the exposure. To infer causality of the association between exposure and outcome, three assumptions have to be fulfilled: (1) GRS must be reproducibly associated with the exposure, (2) must not be associated with confounders and (3) its association with the outcome should pass only through the exposure (figure 1).

The availability of information regarding genetic variants, exposure and outcome for each individual, allowed us to perform individual-level (one-sample) MR analyses.

First, regressions of the GRS on the exposures (YoE and RII_t) were performed to evaluate the strength of the instrumental variables through F-statistic (expected to be higher than 10[24]) and the explained variance (R^2).

Afterwards, the causal estimates were estimated through the two-stage least square (2SLS) method, which consists of a first stage regression of the genetic score on the exposure and of a second stage regression of the predicted values obtained from the first stage on the outcome. Due to the nature of the study (case cohort) and of the outcome, we performed a linear regression of the genetic scores on the exposures only in the subcohort and then Prentice-weighted Cox regression of the predicted values on the incidence of T2D in the overall case cohort. We adjusted both first and second stage of regression for the first 10 principal components of ancestry and genotyping array. Lastly, as for observational analyses, we additionally adjusted for sex, age and country only when YoE was considered as exposure, not with RII_t.

We additionally performed MR analyses considering as outcomes the putative mediators of the association. As represented in figure 2, the presence of mediators does not violate the assumptions of MR,[25] while it allows us to perform further MR analyses considering the same GRS. BMI, daily alcohol consumption, Mediterranean diet score, smoking habits and physical activity were considered as putative mediators. These variables were all collected after completion of participant's highest attained educational level, and thus after the exposure and before the outcome of our study, as a causal sequence requires.

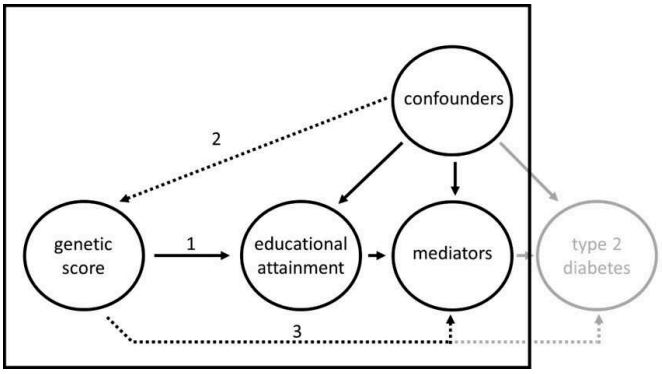

Figure 2 DAG representing the hypothesised connections considered to perform the MR analyses (reported in the rectangle) to investigate the association between educational attainment and putative mediators of the association between educational attainment and T2D. Solid lines represent the associations to be fulfilled, while dotted lines represent the associations which are supposed to be absent to perform MR. The numbers denote the MR assumptions as described in the main text. DAG, directed acyclic graph; MR, Mendelian randomisation; T2D, type 2 diabetes.

Similar to the primary MR analyses, the causal estimates for BMI, daily alcohol consumption and Mediterranean diet score were estimated through the 2SLS method with unaltered first stage (linear regressions), and linear regressions for the second stages, both performed in the subcohort. For dichotomous outcomes (smoking status and physical activity) we conducted the first stage in the group of controls (never smokers or inactive/moderately inactive subcohort individuals) and, as second stage, we estimated OR through logistic regressions in the overall subcohort.

Sensitivity analyses

To investigate one of the biggest threats to MR, that is, horizontal pleiotropy, a set of sensitivity analyses were performed, performing a two-sample MR. Summary statistics for the educational attainment were obtained by the original GWAS,[9] while summary statistics for the outcome (T2D) were computed on this EPIC-Interact case-cohort study.

Analyses were performed by using the Software R (V.4.2.3)[26–28]

RESULTS

Descriptive analysis

Analyses were performed on 12 549 individuals in the random subcohort, and 9568 participants in the case group, with an overlap of 587 individuals (online supplemental figure 1).

Table 1 shows the baseline characteristics of individuals in the EPIC-InterAct case-cohort study.

Observational analysis

The Prentice-weighted Cox models showed a statistically significant inverse association between educational level and the incidence of T2D. A higher educational level, both measured through YoE and RII_t, was associated with a lower risk of developing T2D (HR_{YoE} 0.956, 95% CI 0.951 to 0.961); HR_{RIIt} 0.959, 95% CI 0.954 to 0.964).

Results of the association between YoE or RII_t and possible mediators of the association are reported in online supplemental table 2 and depicted in figure 3.

Table 1 Descriptive analyses for participants in the EPIC-InterAct case-cohort study

		Random subcohort	Total cases
N		11 962	8981
Age, years (mean±SD)		52.25±9.23	55.64±7.62
Sex N (%)	M	4533 (37.9)	4504 (50.2)
	F	7429 (62.1)	4477 (49.8)
Follow-up time, days (mean±SD)		4445.36±734.66	2477.90±1207.49
Years of education (mean±SD)		11.03±5.50	9.64±5.04
Country N (%)	France	318 (2.7)	143 (1.6)
	Italy	1426 (11.9)	1009 (11.2)
	Spain	2354 (19.7)	1591 (17.7)
	UK	899 (7.5)	662 (7.4)
	The Netherlands	1142 (9.5)	594 (6.6)
	Germany	1734 (14.5)	1315 (14.6)
	Sweden	2323 (19.4)	1994 (22.2)
	Denmark	1766 (14.8)	1673 (18.6)
Relative Index of Inequality (mean±SD)		0.47±0.27	0.53±0.26
Relative Index of Inequality transformed (mean±SD)		10.55 (5.42)	9.32 (5.24)
BMI, kg/m^2 (mean±SD)		25.78±4.04	29.66±4.74
Daily alcohol intake, g (mean±SD)		13.52±18.50	14.70±21.60
Mediterranean Diet Score (mean±SD)		8.53±3.13	8.05±3.18
Cambridge Physical Activity Index N (%)	Inactive	2676 (22.5)	2655 (29.8)
	Moderately inactive	3998 (33.7)	2943 (33.0)
	Moderately active	2744 (23.1)	1796 (20.1)
	Active	2454 (20.7)	1529 (17.1)
Smoking status N (%)	Never	5547 (46.6)	3621 (40.5)
	Former	3223 (27.1)	2795 (31.3)
	Smoker	3131 (26.3)	2524 (28.2)

BMI, body mass index; EPIC, European Prospective Investigation into Cancer and Nutrition.

MR analysis on T2D

As shown in table 2, according to both the indices F-statistic and R^2, the GRS appeared to be a good instrumental variable both for RII_t and YoE as exposures (with better values with RII_t). Both the linear regressions of the GRS on the exposures (1° stage) showed a significant and positive association, that is, a correspondence between the increasing in the GRS and the level of education. Similarly, both the causal estimates obtained by the 2° stage showed a protective and significant role of a higher level of education for the development of T2D (HR<1).

The two-sample MR analysis aimed at detecting horizontal pleiotropy suggested that no individual SNP significantly biased our results due to pleiotropic effects. The average pleiotropic effect was near zero and appeared to be well balanced (online supplemental material).

Analysis on potential mediators

MR analyses performed considering as outcomes putative mediators showed that higher educational level may be causally linked to lower BMI and likelihood of smoking, and to higher alcohol consumption and greater adherence to the Mediterranean diet, while no significant results were found for physical activity. Details are reported in figure 3 and online supplemental table 2.

DISCUSSION

Our study investigated the association between educational attainment, as proxy for SEP, and incidence of T2D. Our results suggest that a higher level of YoE is causally associated with a lower risk T2D with an effect approximately equivalent to 15% lower risk for each additional year of education. In an investigation on lifestyle-related factors suggests that a greater number of YoE is causally associated with lower BMI, lower likelihood of smoking, higher alcohol consumption and greater adherence to the Mediterranean diet. These factors may be potential mediators in the association, as their plausibility within the first part of the causal pathway is supported, that is, the exposure appears to influence some potential mediators.

Previous analyses of SEP and T2D

Several researchers have investigated the association between SEP and T2D in observational studies[29 30] providing generally consistent evidence of an association, regardless of which index is used to measure SEP.

In a recent review, Lago-Peñas *et al*[29] reported uniform results among different studies evaluating the association between both childhood and adulthood SEP with incidence of T2D in a later age. Regardless of its measurements, low SEP was consistently associated with a higher risk of developing T2D.

Previous MR analyses

In recent years, the availability of sequencing data in large consortia has allowed researchers to deploy MR approaches to investigate the causal nature of relationships described in observational studies. Several MR analyses have been performed to explore the association between SEP and T2D incidence, with different results.[11] In 2020, two-sample MR analyses[11 12] showed a statistically significant protective effect of additional years of schooling on T2D.[12] Subsequently, Liang *et al*,[13] Zhang *et al*[14] and Davies *et al*[17] confirmed this result.

However, Tillmann *et al*[15] estimated an association with T2D in the anticipated direction, but with a CI including the null value, and Ne-Ek *et al*[16] reported inconsistent findings between the cohort study and MR analyses: even though observational evidence highlighted that individuals in the low education level were at increased risk of developing T2D compared with the high education group, MR analysis suggested null associations.

Potential mechanisms

Traditional observational studies have estimated that about 30%–45% of the association between SEP and T2D may be explained by differences in behavioural risk factors, such as obesity, smoking and alcohol consumption, and in biological risk markers, such as systolic blood pressure and triglycerides.[31 32] A two-step two-sample MR analysis,[14] using genetic variants associated with putative mediators, has estimated a higher proportion, and accounted for about 80% of the total effect, explained by BMI, smoking, watching television, and systolic and diastolic blood pressure. Evidence obtained through these different study designs[14 31 32] suggests that conventional risk factors may explain only a fraction of the total effect, suggesting the need to investigate other factors that may mediate the relationship between education and T2D risk. These could include the fact

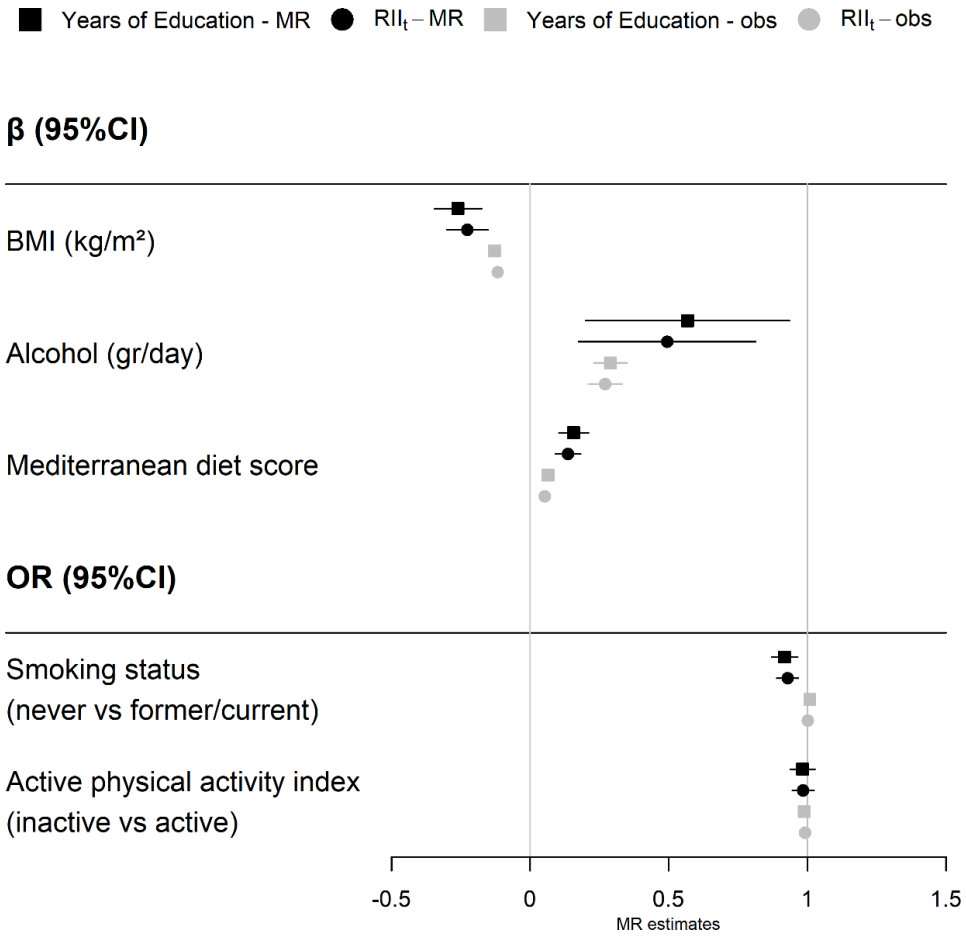

Figure 3 Results of the observational and MR analyses investigating the association between educational attainment (EA) and potential mediator of the association between EA (measured as YoE and RII$_t$) and T2D in the EPIC-InterAct sub-cohort. BMI, body mass index; EPIC, European Prospective Investigation into Cancer and Nutrition; MR, Mendelian randomisation; RII$_t$, transformed Relative Index of Inequality; T2D, type 2 diabetes.

that education promotes better health service utilisation, from higher health literacy (ie, adherence to health behaviour recommendations, interpretation of symptoms, healthcare seeking and communication) to fewer financial barriers to care.[14]

Furthermore, several hypotheses have been put forward in order to understand if the cumulative burden caused by chronic challenging conditions may explain how SEP gets biologically embedded. People with a low SEP may experience conditions of stress more often compared with their high SEP counterparts, independently of behavioural risk factors.[33] In addition to systolic and diastolic blood pressure, other altered multisystem responses to stress, such as heart rate, total cholesterol, salivary cortisol and plasma IL-6 levels, could be measured and considered as putative mediators.[34]

Table 2 Results of the MR analyses on the association between educational attainment and type 2 diabetes in the EPIC-InterAct case-cohort study

		F-statistic	R^2	1° stage (β (95%CI))	2° stage (HR (95%CI))
GRS	YoE	258.1	2.0%	3.043 (2.720 to 3.366)	0.834 (0.805 to 0.864) (0.812 to 0.864)*
	RII$_t$	381.0	2.9%	3.486 (3.134 to 3.838)	0.854 (0.829 to 0.880) (0.826 to 0.880)*

1° stage is the regression of the GRS on the exposure (YoE or RII$_t$), 2° stage is the regression of the predicted values in the 1° stage on the outcome (type 2 diabetes).
*CI obtained by the bootstrap method to take into account the uncertainty introduced by the two-stage process.
EPIC, European Prospective Investigation into Cancer and Nutrition; GRS, Genetic Risk Score; MR, Mendelian randomisation; RII$_t$, transformed Relative Index of Inequality; YoE, years of education.

Strength and limitations

Our study adds high-quality evidence to similar results obtained by different epidemiological study designs, reinforcing the plausibility of the reported associations. The strengths of this study lie in a robust and well-powered European case-cohort design, with the availability of individual level data on each variable involved in the framework and in the MR approach. This has made it possible to avoid common biases in observational studies.[7]

We avoided the threat of the so-called winner's curse phenomenon, that may cause an overestimation of the true causal effect,[35] by selecting the instruments for the exposure from external GWAS.

Furthermore, our one-sample MR analysis offers the possibility of comparing the results obtained by two-sample MR analyses. Genetic variants are likely to explain a small proportion of SEP variance and so be weak predictors of it. When weak instruments are used, one-sample and two-sample MR analyses may be biased in different directions. One-sample MR analyses will give results biased towards the confounded regression analysis results, while two-sample MR analyses will return results biased towards the null.[36] Our one-sample MR analysis returned results similar to the already cited previous two-sample MR analyses,

with the exception of,[16] in which null results may be explained by weaker instruments for the exposure.

However, specific attention should be paid when applying the MR approach to complex and distal phenotypes, such as education attainment, where the effects of genetic variants may operate through composite biological, behavioural and social pathways.[37] Indeed, genetic variants may affect a phenotype different from the exposure of interest which may not be on the causal pathway between the exposure and the outcome, resulting in a phenomenon called horizontal pleiotropy, which violates the third assumption required by MR. To our knowledge, the variants included in the GRS were not associated with any factors that have these characteristics and a two-sample MR, with specific sensitivity analyses to detect horizontal pleiotropy, showed no evidence of pleiotropic effects.

When an offspring's phenotype is influenced not just through the inheritance of the genotype, but also by parents' phenotype, the so-called dynastic effect may occur.[38] This happens when the parents' phenotype affects the environment in which their offspring grows with the consequence that offspring phenotype is driven by environmental inheritance, even when those offspring do not inherit the genetic variants associated with the phenotype. Dynastic effect can cause bias and false positive results in MR studies in social epidemiology. However, only cohort studies that incorporate genetic information from sibling samples, parents and offspring would allow for the investigation and reduction of these possible biases.[39]

Lastly, our analyses exploit GWAS performed on Western European populations, which reduces the generalisability of the results to populations with different genetic ancestry.[40]

CONCLUSION

The results from this study contribute to our understanding of the association between educational attainment and the incidence of T2D. Educational attainment is influenced by a wide range of factors and it is not solely determined by genetics. The GRS served as a weak instrument for educational attainment but allowed us to use genetic findings to enhance the evidence on this significant topic. MR enables causal estimation by simulating an experimental design within an observational framework, minimising the potential effects of residual confounding or reverse causation, but should always be reinforced together with findings from other epidemiological study designs. No single statistical method can definitively answer causal questions but combining evidence obtained from diverse methodologies can help to validate research hypotheses.

Author affiliations
[1]Department of Clinical and Biological Sciences, University of Turin, Orbassano, Italy
[2]Department of Translational Medicine, University of Eastern Piedmont, Novara, Italy
[3]Department of Health Sciences, University of Eastern Piedmont, Novara, Italy
[4]Julius Center for Health Sciences and Primary Care, University Medical Center, Utrecht, The Netherlands
[5]Unit of Nutrition and Cancer, Cancer Epidemiology Research Programme, Catalan Institute of Oncology, Bellvitge Biomedical Research Institute, Barcelona, Spain
[6]International Agency for Research on Cancer, Lyon, France
[7]Université Paris-Saclay, UVSQ, Inserm, Gustave Roussy, CESP, 94805, Villejuif, France
[8]Epidemiology and Prevention Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milano, Italy
[9]Clinical Epidemiology Unit, Institute for Cancer Research, Prevention and Clinical Network (ISPRO), Florence, Italy
[10]School of Medicine, Federico II University, Naples, Italy
[11]Department of Epidemiology and Biostatistics, Imperial College London, London, UK
[12]Department of Inflammation Biology, King's College London, London, UK
[13]German Institute of Human Nutrition Potsdam-Rehbruecke, Nuthetal, Germany
[14]German Center for Diabetes Research, Neuherberg, Germany
[15]Institute of Nutritional Science, University of Potsdam, Nuthetal, Germany
[16]German Cancer Research Center, Heidelberg, Germany
[17]Department of Public Health, Aarhus University, Aarhus C, Denmark
[18]Danish Cancer Institute, Danish Cancer Society, Copenhagen, Denmark
[19]Department for Clinical Oncology & Palliative Care, Zealand University Hospital, Naestved, Denmark
[20]Department of Epidemiology, Murcia Regional Health Council, Murcia, Spain
[21]CIBERESP, Madrid, Spain
[22]Research Group on Demography and Health, National Faculty of Public Health, University of Antioquia, Medellin, Colombia
[23]Public Health Institute of Navarra, Pamplona, Spain
[24]Ministry of Health of the Basque Government, San Sebastián, Spain
[25]BioGipuzkoa Health Research Institute, San Sebastián, Spain
[26]Andalusian School of Public Health, Granada, Spain
[27]Instituto de Investigación Biosanitaria ibs.GRANADA, Granada, Spain
[28]MRC Epidemiology, University of Cambridge, Cambridge, UK

Correction notice This article has been updated since it first published. It is now open access.

Acknowledgements The National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands, and AIRE ONLUS for their contribution and ongoing support to the EPIC Study. The coordination of EPIC is financially supported by the International Agency for Research on Cancer (IARC) and also by the Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, which has additional infrastructure support provided by the NIHR Imperial Biomedical Research Centre (BRC). The national cohorts are supported by Danish Cancer Society (Denmark); Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l'Education 17 Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM) (France); German Cancer Aid, German Cancer Research Center (DKFZ), German Institute of Human Nutrition PotsdamRehbruecke (DIfE), Federal Ministry of Education and Research (BMBF) (Germany); Associazione Italiana per la Ricerca sul Cancro (AIRC), Compagnia di San Paolo and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (The Netherlands); Health Research Fund (FIS), Instituto de Salud Carlos III (ISCIII), Regional Governments of Andalucía, Asturias, the Basque Country, Murcia and Navarra, and the Catalan Institute of Oncology (ICO) (Spain); Swedish Cancer Society, Swedish Research Council, and County Councils of Skåne and Västerbotten (Sweden); Cancer Research UK (C864/A14136 to EPIC-Norfolk; C8221/A29017 to EPIC-Oxford), Medical Research Council (MR/N003284/1, MC-UU_12015/1 and MC_UU_00006/1 to EPIC-Norfolk; MR/M012190/1 to EPIC-Oxford) (UK).

Contributors FR, CS and NW conceived and designed the analysis. AM, CDG and MF performed the analysis. CG aided in enhancing the analysis. AM, FR and CS provided a preliminary interpretation of the results and drafted the manuscript. AM, CS, CG, CDG, MF, YTvdS, RZ-R, EW, CD, AE, TT, CA, BB, SP, PV, SC, MBS, VK, RB, CCD, SOD, SMC-Y, CM-I, PAE, MJS, NGF, NW and FR contributed to the interpretation of the results, revised the draft and approved the final version of the manuscript. CS is the guarantor.

Funding RZ-R was supported by the 'Miguel Servet II' (CPII20/00009) program from the Instituto de Salud Carlos III (cofunded by European Social Fund (ESF) investing in your future). The EPIC-InterAct project was funded by the European Union FP6 programme (grant no. LSHM_CT_2006_037197).

Disclaimer Where authors are identified as personnel of the International Agency for Research on Cancer/WHO, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer/WHO. The authors have no relevant financial or non-financial interests to disclose.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and the EPIC-InterAct study was approved by the local ethics committee in the participating countries and the Internal Review Board of the International Agency for Research on Cancer. All participants gave written informed consent. The EPIC-Interact study was led by the Medical Research Council Epidemiology Unit at the University of Cambridge. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. Raw data cannot be made freely available due to restrictions imposed by the ethical committees which do not allow open/public sharing of data on individuals. However, aggregated data are available on request. Requests should be sent to the corresponding author.

ORCID iDs
Alessandra Macciotta http://orcid.org/0000-0002-0287-7979
Alexis Elbaz http://orcid.org/0000-0001-9724-5490
Paolo Vineis http://orcid.org/0000-0001-8935-4566
Fulvio Ricceri http://orcid.org/0000-0001-8749-9737

REFERENCES

1 Khan MAB, Hashim MJ, King JK, *et al*. Epidemiology of Type 2 Diabetes - Global Burden of Disease and Forecasted Trends. *J Epidemiol Glob Health* 2020;10:107–11.
2 Seuring T, Archangelidi O, Suhrcke M. The Economic Costs of Type 2 Diabetes: A Global Systematic Review. *Pharmacoeconomics* 2015;33:811–31.
3 Agardh E, Allebeck P, Hallqvist J, *et al*. Type 2 diabetes incidence and socio-economic position: a systematic review and meta-analysis. *Int J Epidemiol* 2011;40:804–18.
4 Sacerdote C, Ricceri F, Rolandsson O, *et al*. Lower educational level is a predictor of incident type 2 diabetes in European countries: the EPIC-InterAct study. *Int J Epidemiol* 2012;41:1162–73.
5 Brown AF, Ettner SL, Piette J, *et al*. Socioeconomic position and health among persons with diabetes mellitus: a conceptual framework and review of the literature. *Epidemiol Rev* 2004;26:63–77.
6 Altman DG. Randomisation. *BMJ* 1991;302:1481–2.
7 Didelez V, Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. *Stat Methods Med Res* 2007;16:309–30.
8 Okbay A, Beauchamp JP, Fontana MA, *et al*. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* 2016;533:539–42.
9 Lee JJ, 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;50:1112–21.
10 Qi Y, Koster A, van Boxtel M, *et al*. Adulthood Socioeconomic Position and Type 2 Diabetes Mellitus-A Comparison of Education, Occupation, Income, and Material Deprivation: The Maastricht Study. *Int J Environ Res Public Health* 2019;16:1435.
11 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.
12 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.
13 Liang J, Cai H, Liang G, *et al*. Educational attainment protects against type 2 diabetes independently of cognitive performance: a Mendelian randomization study. *Acta Diabetol* 2021;58:567–74.
14 Zhang J, Chen Z, Pärna K, *et al*. Mediators of the association between educational attainment and type 2 diabetes mellitus: a two-step multivariable Mendelian randomisation study. *Diabetologia* 2022;65:1364–74.
15 Tillmann T, Vaucher J, Okbay A, *et al*. Education and coronary heart disease: mendelian randomisation study. *BMJ* 2017;358:j3542.
16 Na-Ek N, Srithong J, Aonkhum A, *et al*. Educational level as a cause of type 2 diabetes mellitus: Caution from triangulation of observational and genetic evidence. *Acta Diabetol* 2022;59:127–35.
17 Davies NM, Dickson M, Davey Smith G, *et al*. The causal effects of education on adult health, mortality and income: evidence from Mendelian randomization and the raising of the school leaving age. *Int J Epidemiol* 2023;52:1878–86.
18 Riboli E, Hunt KJ, Slimani N, *et al*. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr* 2002;5:1113–24.
19 Forouhi NG, Wareham NJ. The EPIC-InterAct Study: A Study of the Interplay between Genetic and Lifestyle Behavioral Factors on the Risk of Type 2 Diabetes in European Populations. *Curr Nutr Rep* 2014;3:355–63.
20 Burgess S, Thompson SG. Use of allele scores as instrumental variables for Mendelian randomization. *Int J Epidemiol* 2013;42:1134–44.
21 Ricceri F, Sacerdote C, Giraudo MT, *et al*. The Association between Educational Level and Cardiovascular and Cerebrovascular Diseases within the EPICOR Study: New Evidence for an Old Inequality Problem. *PLoS One* 2016;11:e0164130.
22 Salas-Salvadó J, Bulló M, Babio N, *et al*. Reduction in the incidence of type 2 diabetes with the Mediterranean diet: results of the PREDIMED-Reus nutrition intervention randomized trial. *Diabetes Care* 2011;34:14–9.
23 InterAct Consortium, Peters T, Brage S, *et al*. Validity of a short questionnaire to assess physical activity in 10 European countries. *Eur J Epidemiol* 2012;27:15–25.
24 Burgess S, Thompson SG, CRP CHD Genetics Collaboration. Avoiding bias from weak instruments in Mendelian randomization studies. *Int J Epidemiol* 2011;40:755–64.
25 Solovieff N, Cotsapas C, Lee PH, *et al*. Pleiotropy in complex traits: challenges and strategies. *Nat Rev Genet* 2013;14:483–95.
26 Package 'boot'. 2024.
27 Package 'survival'. 2024. Available: https://github.com/therneau/survival
28 Hemani G, Zheng J, Elsworth B, *et al*. The MR-Base platform supports systematic causal inference across the human phenome. *Elife* 2018;7:e34408.
29 Lago-Peñas S, Rivera B, Cantarero D, *et al*. The impact of socioeconomic position on non-communicable diseases: what do we know about it? *Perspect Public Health* 2021;141:158–76.
30 Smith BT, Lynch JW, Fox CS, *et al*. Life-course socioeconomic position and type 2 diabetes mellitus: The Framingham Offspring Study. *Am J Epidemiol* 2011;173:438–47.
31 Stringhini S, Tabak AG, Akbaraly TN, *et al*. Contribution of modifiable risk factors to social inequalities in type 2 diabetes: prospective Whitehall II cohort study. *BMJ* 2012;345:e5452.
32 Steele CJ, Schöttker B, Marshall AH, *et al*. Education achievement and type 2 diabetes-what mediates the relationship in older adults? Data from the ESTHER study: a population-based cohort study. *BMJ Open* 2017;7:e013569.
33 McCrory C, Fiorito G, Ni Cheallaigh C, *et al*. How does socio-economic position (SEP) get biologically embedded? A comparison of allostatic load and the epigenetic clock(s). *Psychoneuroendocrinology* 2019;104:64–73.
34 Guidi J, Lucente M, Sonino N, *et al*. Allostatic Load and Its Impact on Health: A Systematic Review. *Psychother Psychosom* 2021;90:11–27.
35 Taylor AE, Davies NM, Ware JJ, *et al*. Mendelian randomization in health research: using appropriate genetic variants and avoiding biased estimates. *Econ Hum Biol* 2014;13:99–106.
36 Lawlor DA. Commentary: Two-sample Mendelian randomization: opportunities and challenges. *Int J Epidemiol* 2016;45:908–15.
37 Munafò M, Davies NM, Davey Smith G. Can genetics reveal the causes and consequences of educational attainment? *J R Stat Soc Ser A Stat Soc* 2020;183:681–8.
38 Kong A, Thorleifsson G, Frigge ML, *et al*. The nature of nurture: Effects of parental genotypes. *Science* 2018;359:424–8.
39 Brumpton B, Sanderson E, Heilbron K, *et al*. Avoiding dynastic, assortative mating, and population stratification biases in Mendelian randomization through within-family analyses. *Nat Commun* 2020;11:3519:3519:.
40 Peter BM, Petkova D, Novembre J. Genetic Landscapes Reveal How Human Genetic Diversity Aligns with Geography. *Mol Biol Evol* 2020;37:943–51.