Association and mediation between educational attainment and respiratory diseases: a Mendelian randomization study

Guohui Lan[1†], Mengying Xie[2†], Jieli Lan[3†], Zelin Huang[1†], Xiaowei Xie[4], Mengdan Liang[1], Zhehui Chen[1], Xiannuan Jiang[1], Xiaoli Lu[1], Xiaoying Ye[1], Tingting Xu[1], Yiming Zeng[5*] and Xiaoxu Xie[1,3,5*]

Abstract

Background Respiratory diseases are a major health burden, and educational inequalities may influence disease prevalence. We aim to evaluate the causal link between educational attainment and respiratory disease, and to determine the mediating influence of several known modifiable risk factors.

Methods We conducted a two-step, two-sample Mendelian randomization (MR) analysis using summary statistics from genome-wide association studies (GWAS) and single nucleotide polymorphisms (SNPs) as instrumental variables for educational attainment and respiratory diseases. Additionally, we performed a multivariable MR analysis to estimate the direct causal effect of each exposure variable included in the analysis on the outcome, conditional on the other exposure variables included in the model. The mediating roles of body mass index (BMI), physical activity, and smoking were also assessed.

Findings MR analyses provide evidence of genetically predicted educational attainment on the risk of FEV1 ($\beta = 0.10$, 95% CI 0.06, 0.14), FVC ($\beta = 0.12$, 95% CI 0.07, 0.16), FEV1/FVC ($\beta = -0.005$, 95% CI -0.05, 0.04), lung cancer (OR = 0.54, 95% CI 0.45, 0.65) and asthma (OR = 0.86, 95% CI 0.78, 0.94). Multivariable MR dicated the effect of educational attainment on FEV1 ($\beta = 0.10$, 95% CI 0.04, 0.16), FVC ($\beta = 0.07$, 95% CI 0.01, 0.12), FEV1/FVC ($\beta = 0.07$, 95% CI 0.01, 0.01), lung cancer (OR = 0.55, 95% CI 0.42, 0.71) and asthma (OR = 0.88, 95% CI 0.78, 0.99) persisted after adjusting BMI and cigarettes per day. Of the 23 potential risk factors, BMI, smoking may partially mediate the relationship between education and lung disease.

Conclusion High levels of educational attainment have a potential causal protective effect on respiratory diseases. Reducing smoking and adiposity may be a target for the prevention of respiratory diseases attributable to low educational attainment.

Keywords Education, Lung function, Lung cancer, Asthma, Mendelian randomization

†Guohui Lan, Mengying Xie, Jieli Lan and Zelin Huang have contributed equally to this work.

*Correspondence:
Yiming Zeng
zeng_yi_ming@126.com
Xiaoxu Xie
xxx@fjmu.edu.cn
Full list of author information is available at the end of the article

Lan *et al. Respiratory Research*   (2024) 25:115

Page 2 of 11

Key messages

What is already known on this topic

Several observational studies have revealed that people with a higher education attainment is associated with a lower risk of developing respiratory diseases. However, observational studies are susceptible to reverse causation and confounding factors. Also, the role of genetic factors in the study remains unknown.

What this study adds

In this study, by leveraging data from the recently published genome-wide association studies, we found a significant genetic correlation between educational attainment and respiratory disease. We further confirmed that the causal relationship between educational attainment and respiratory disease is partially mediated by smoking and obesity.

How this study might affect research, practice or policy

Our study highlights the importance of early detection and prevention of respiratory disease, including lung function, lung cancer and asthma, amongst low education group. Moreover, our findings might provide new understandings for the mechanisms associated with educational attainment and respiratory disease.

Introduction

Deaths from chronic respiratory diseases constituted 7% of all deaths globally in 2019, with the prevalent diseases including chronic obstructive pulmonary disease (COPD), asthma, and lung cancer [1]. Identifying potential risk factors is crucial to safeguarding public health and preventing the emergence of diseases. Lung function is an important predictor of quality of life and longevity [2].

Socioeconomic disparities in health have been documented. Individuals with lower socioeconomic status have higher mortality and morbidity risks compared to individuals with higher socioeconomic status. There has been research into the impact of socioeconomic factors on health outcomes [3–7]. Among many socioeconomic indicators, educational attainment (EA) has been identified as a social determinant of health through various mechanisms, such as neurodevelopment, health behavior, and health literacy [8].

Several studies have examined the association between EA and respiratory diseases. Previous studies have employed cross-sectional designs to investigate the complex relationship between EA and lung function [9] and lung cancer [10]. There have been several studies that have examined the effects of education on pulmonary health, and they have also identified potential mechanisms or mediators that may explain these effects. The finding suggest that this association may be mediated by some modifiable factors related to both exposure and outcome, such as BMI, physical activity and smoking. Nevertheless, these studies were observational in nature and thus prone to methodological limitations, including confounding and reverse causality, as well as failure to consider mediating factors. Therefore, the relationship between EA and respiratory disease, as well as lung function is unclear.

Mendelian randomization (MR) is a method for inferring causal relationships based on genetics, utilizing single-nucleotide polymorphism (SNP) as a surrogate of exposure, evaluating observed data and correlating causal relationships through a statistical relationship between genotypes and phenotypes [11]. Under several assumptions, an MR study should produce results which avoid the potential biases associated with observational studies, such as confounding, reverse causation, and measurement errors, which are common in observational studies. Multivariable Mendelian randomization (MVMR) is a rapidly evolving analytical method that estimates the effect of each exposure variable on the model results while also considering the effects of other exposure variables that may affect the model results [12]. This method is based on Mendelian inheritance law, randomly grouping multiple variables simultaneously to create a random distribution of variables between the groups, thereby enhancing the reliability and accuracy of the experiment.

The correlation between EA and asthma, COPD, and lung cancer has been reported in previous studies [9, 10, 13], the causal mechanism is not clear and these did not assess mediation by modifiable factors. In addition, recent GWAS on compared with the previously reported GWAS, a newly published GWAS for EA comes from a large sample of population data, and the results are more accurate. Therefore, using the latest GWAS data, it is possible to update the results of studies on the relationship between EA and respiratory diseases, in order to better understand this association. Potential confounding factors were also included in the MVMR analysis to control for their effects and obtain more accurate estimates of the direct causal effect of each exposure on the outcome.

In this report, the MR method was used to assess the causal association between EA and lung function, asthma and lung cancer, and a two-step Mendelian randomization was used to assess its mediated proportion in association for 23 potential mediating factors. Ultimately these causal conclusions will support the development of prevention policies.

Lan *et al. Respiratory Research*   (2024) 25:115

Page 3 of 11

Methods

Study design

In Mendelian randomization research, genetic information is usually used as an instrumental variable (IV) due to their random distribution in humans and robust associations with the exposure and outcome variables being investigated. EA was assessed causally associated with lung cancer, FEV1, FVC, FEV1/FVC and asthma using two-step Mendelian randomization analysis. All GWAS summary statistics were gathered from a public GWAS website (https://gwas.mrcieu.ac.uk/) for the purposes of these analyses. Data summarized from GWAS are presented in Table 1.

Education attainment

The genetic instruments for EA were selected from a meta-analysis comprising 71 GWAS discovery cohorts that included 1, 131, 881 European ancestral participants. To facilitate the classification and conversion of educational levels into standardized units for better cross-country and cross-regional comparisons, the International Standard Classification of Education (ISCED) 2011 was employed, utilizing 4.2 years of education as the unit within the educational systems of the UK and the US. In a study with the identifier ieu-a-1239, conducted on a sample of more than 1.1 million individuals, a genetic association analysis of EA was performed. This analysis identified 1271 independent SNPs that exhibited significant associations with EA [14].

Outcome-respiratory disease

The outcomes used in this study were lung function indicators and related lung diseases (lung cancer and asthma).

Lung function

We selected respiratory function indicators to assess lung function in 400,102 European ancestry individuals. We identified 139 new signals related to lung function, including forced expiratory volume in one second (FEV1), lower forced vital capacity (FVC), and the FEV1-to-FVC ratio [15]. ID: ebi-a-GCST007431 (FEV1/FVC), ebi-a-GCST007429 (FVC), ebi-a-GCST007432 (FEV1).

Lung cancer

The International Lung Cancer Consortium (ILCCO) conducted a GWAS analysis on lung cancer and identified 259 SNPs (with a significance level of $P < 5 \times 10^{-8}$) in a study involving 11,348 lung cancer cases and 15,861 controls [16]. ID: ieu-a-966.

Asthma

Valette et al.[17] used genetic instruments from the UK Biobank in a study that employed a broad definition of asthma. The study included 56,167 asthma cases and 352,255 controls. ID: ebi-a-GCST90014325.

Mediators

Based on our review of the literature, we selected 23 candidate mediators of modifiable risk factors (please refer to Additional file 1: Fig. S1 in the supplementary materials for an overview of the process of identifying the candidate mediators). The mediators involved in the relationship between EA and respiratory disease were selected based on the following criteria for inclusion in the analysis: (1) Exposure and mediating factors are related in a causal way; (2) There was an association between mediating factors and outcomes, whether or not exposure factors were corrected. Ultimately, we identified three risk factors that met the criteria, including BMI [18], physical activity [19] and cigarettes per day [20], were included

Table 1 Overview of GWAS data used in multivariable Mendelian randomization (MVMR)

Phenotype	Number of participants	Number of SNP	Consortium	Author	PubMed ID	GWAS ID
Year of schooling	766,345	10,101,242	SSGAC	Lee et al.	30038396	ieu-a-1239
FEV1	321,047	19,641,887	–	Shrine N	30804560	ebi-a-GCST007432
FVC						ebi-a-GCST007429
FEV1/FVC						ebi-a-GCST007431
Lung cancer	27,209	8,945,893	ILCCO	Wang Y et al.	24880342	ieu-a-966
Asthma	408,442	34,551,291	–	Valette K et al.	34103634	ebi-a-GCST90014325
BMI	681,275	2,336,260	GIANT	Loic Yengo et al.	30124842	ieu-b-40
physical activity	215,127	15,538,177	–	Tyrrell J et al.	33563987	ebi-a-GCST90012791
Cigarettes per Day	337,334	11,913,712	GSCAN	Liu, M et al.	30643251	ieu-b-25

SSGAC Social Science Genetic Association Consortium, *ILCCO* International Lung Cancer Consortium, *GIANT* Genetic Investigation of Anthropometric Traits, *GSCAN* GWAS & Sequencing Consortium of Alcohol and Nicotine use, *EA* educational attainment, *BMI* body mass index, *FEV1* forced expiratory volume in one second, *FVC* forced vital capacity, *FEV1/FVC* forced expiratory volume in one second / forced vital capacity

Lan *et al. Respiratory Research*   (2024) 25:115

Page 4 of 11

in the mediation analysis to assess the role of mediation between EA and lung function, lung cancer, or asthma. ID: ieu-b-40 (BMI), ebi-a-GCST90012791 (physical activity), ieu-b-25 (cigarettes per day).

SNP selection

To conduct Mendelian randomization, we selected the instrumental variables (IVs) as follows. Firstly, we selected SNPs that were significantly associated with educational attainment for each MR analysis, excluding genetic instruments with P values greater than 5×10^{-8} in relation to the exposure. Secondly, as part of the MR analysis, we utilized independent SNPs as genetic instruments when genetic associations were identified for both the exposure and the outcome of interest. Then, the clumping process ($r^2 < 0.001$ within 10,000-kb windows) was employed to determine whether the included SNPs are in linkage disequilibrium (LD). If no SNPs related to exposure were identified in the results, we did not utilize proxy SNPs. Finally, to ensure that there was no direct correlation between the instrumental variables used in the analysis and the outcome, excluding genetic instruments with P values $< 5 \times 10^{-8}$ in relation to the outcome.

Statistical analysis

Based on three critical assumptions, the MR method was developed: (1) The genetic variation must be closely related to the exposure in the MR analysis; (2) Genetic variation cannot be associated with confounding factors between exposure and outcome; (3) Exposure must be the mechanism through which genetic variables influence outcomes [21].

To assess whether potential mediators mediate between exposure and outcome, a two-step Mendelian randomization was used to assess the effect. The first step involved estimating the effect sizes of the exposure on lung function, lung cancer, asthma and mediators respectively. We use IVW as our primary method, which is characterized by regression without considering the presence of intercept terms and fitting with the reciprocal of the outcome variance as a weighting factor [22]. Additionally, we used MR-PRESSO, MR-Egger, and weighted median tests to estimate the effects. Subsequently, MVMR was used to determine the effect of each mediator on the outcome, taking into account how each instrument was genetically influenced [23].

Direct and indirect effects are both part of the total effect. Direct effect refers to the impact of the exposure factor on the outcome, independent of intermediary variables. Indirect effect, on the other hand, refers to the impact of the exposure factor on the outcome through intermediary variables [24]. The overall effect of EA on outcome was thus decomposed into two distinct components: (i) the indirect effect through each mediator individually, indicating the influence of education As a primary method for testing whether a mediated effect was present and its magnitude, we used the Sobel test (a×b in Fig. 1), and (ii) the direct effect of education on outcome after adjusting for each mediator (c' in Fig. 1) [25]. By using this statistical technique, we can explore complex relationships between variables and understand how intermediary variables impact exposure-outcome relationships.

To derive the indirect effect of combining multiple mediations, the difference method (c–c') is used, where c' indicates that multiple mediating factors are adjusted in the MVMR model. The delta method is used to calculate the confidence interval when the indirect effect is divided by the total effect (RMediation (shinyapps.io)), the proportion of the mediating effect can be quantified for one mediator or a combination of mediations. For each genetic instruments, we set $P < 5 \times 10^{-8}$ to selected genome-wide significant SNPs. To address the issue of linkage disequilibrium, we applied LD thresholds pairwise from the original GWAS for each mediator, with SNPs for each mediator adhering to an LD cut-off of $r^2 < 0.01$ within a window of 1 MB.

Sensitivity analysis

UVMR's IVW method can be examined for its robustness using two methods. Weighted medians are used in
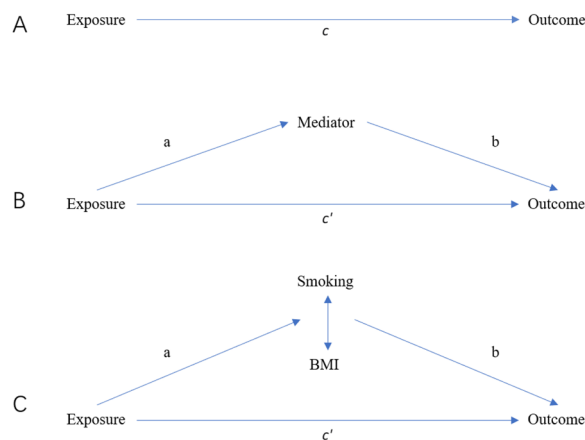

Fig. 1 Diagrams illustrating associations examined in this study. **A** The total effect of exposure on outcome, c, was derived using univariable MR. **B** The total effect was decomposed into: (i) indirect effect using a two-step MR (where *a* is the total effect of exposure on mediator, *b* is the effect of mediator on outcome adjusting for exposure and the mediating effect is calculated using the product method (*a*×*b*)); (ii) direct effect (*c'*=*c* – *a*×*b*). **C** For mediation by both smoking and BMI combined (arrows represent their bidirectional causal relationship), the indirect effect was derived using the difference method (*c*–*c*). Proportion mediated was the indirect effect divided by the total effect

Lan *et al. Respiratory Research*   (2024) 25:115

Page 5 of 11

UVMR as well as Egger methods in MR Egger to assess the robustness of the IVW method, and Egger methods are used in MVMR to assess the robustness of the MVMR-IVW method. The MR-Egger method can determine whether horizontal pleiotropism exists in the instrumental variable to prevent violating the instrumental variable assumption. In addition, the Cochran's Q test is often used as an indicator of heterogeneity in meta-analysis, with a P-value less than 0.05 indicating the presence of heterogeneity in the study. An assessment of the strength of the genetic instrumental variables used in the study was conducted by using conditional F-statistics. A commonly used threshold for an "acceptable" F-statistic is 10, indicating that the instruments explain at least 10 times as much variance as the residual variance. However, this threshold may vary depending on the study design and sample size. In addition, we performed a "leave-one-out" sensitivity assessment to determine whether or not a certain SNP had too much influence on the results, and these SNPs were excluded from the MR analysis. Only when the IVW estimate agrees with at least one sensitivity analysis in direction and statistical significance, and there is no evidence of pleiotropy, is it considered to have a causal association.

The MR analyses were all performed using R (version 4.0.2) with the "TwoSampleMR" and "MRPRESSO" R package [26, 27].

Patient and public involvement
The patient and public were not involved in the design or reporting of this study.

Result

Effect of education attainment on lung function, lung cancer and asthma
The results of analyses found that increased genetically predicted EA was significantly related to enhanced FEV1 ($\beta = 0.10$, 95% CI 0.06, 0.14), improved FVC ($\beta = 0.12$, 95% CI 0.07, 0.16), and a less favorable FEV1/FVC ratio ($\beta = -0.005$, 95% CI − 0.05, 0.04). Furthermore, this heightened EA was also associated with a reduced risk of lung cancer (OR = 0.54, 95% CI 0.45, 0.65) and asthma (OR = 0.86, 95% CI 0.78, 0.94) (Fig. 2).
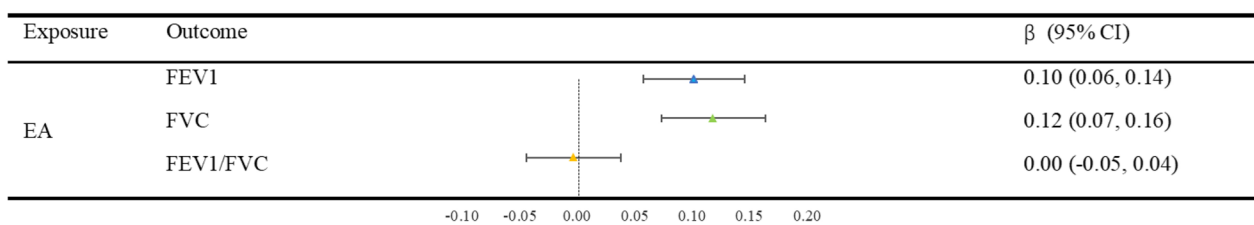
Effect of education attainment on mediators
Table 2 shows the impact of education predicted by genetics on various mediators. A UVMR analysis revealed that for each extra 1-SD year of education are associated with lower BMI (IVW = − 0.16, 95% CI − 0.22, − 0.10), fewer cigarettes smoked per day (IVW = − 0.32, 95% CI − 0.40, − 0.24), and higher physical activity levels (IVW = 0.20, 95% CI 0.16, 0.23).

Effect of mediators on lung function, lung cancer and asthma after adjusting education attainment
According to Fig. 3, each mediator significantly predicted lung function and lung cancer after adjusting for EA. In this study, we excluded physical activity from our analysis because there was only one SNP available, which would lead to a large bias in the results. In the MVMR results, a 1-SD increase in BMI was associated with an increased risk of FEV1/FVC ($\beta = 0.11$, 95% CI 0.09, 0.14); lung cancer (OR = 1.12, 95% CI 0.98, 1.28); asthma (OR = 1.15, 95% CI 1.08, 1.22), and a 1-SD increase in

A. Educational attainment on lung function

Exposure	Outcome	β (95% CI)
EA	FEV1	0.10 (0.06, 0.14)
	FVC	0.12 (0.07, 0.16)
	FEV1/FVC	0.00 (-0.05, 0.04)

-0.10   -0.05   0.00   0.05   0.10   0.15   0.20

B. Educational attainment on lung cancer and asthma

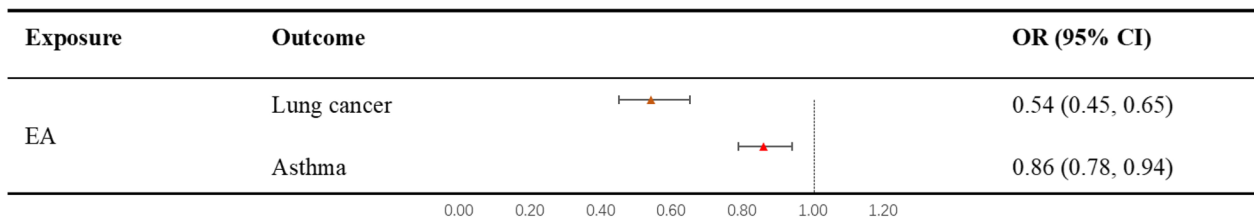| Exposure	Outcome	OR (95% CI)
EA	Lung cancer	0.54 (0.45, 0.65)
	Asthma	0.86 (0.78, 0.94)

0.00   0.20   0.40   0.60   0.80   1.00   1.20

Fig. 2 MR-estimated effects of educational attainment on each outcome separately, presented as β/OR with 95% CI. *EA* educational attainment, *FEV1* forced expiratory volume in one second, *FVC* forced vital capacity, *FEV1/FVC* forced expiratory volume in one second / forced vital capacity

Lan *et al. Respiratory Research*   (2024) 25:115

Page 6 of 11

Table 2 Mendelian randomization analysis of the effect of educational attainment on mediators

Exposure	Outcome	Method	N of SNPs	β	SE	P
EA	BMI	MR Egger	106	− 0.1879	0.1476	0.2059
		Weighted Median	106	− 0.1601	0.0283	1.5467e−8
		IVW	106	− 0.1604	0.0314	3.3414e−07
	Physical activity	MR Egger	263	0.1425	0.0740	5.5370e−02
		Weighted Median	263	0.2020	0.0244	1.2701e−16
		IVW	263	0.1953	0.0186	8.2055e−26
	Cigarettes per day	MR Egger	242	− 0.1910	0.1619	2.3935e−01
		Weighted Median	242	− 0.2816	0.0500	1.7228e−08
		IVW	242	− 0.3218	0.0414	7.2237e−15

EA educational attainment, *BMI* body mass index, *IVW* inverse variance weighted, *MR* Mendelian randomization

genetically predicted cigarettes per day was associated with a higher risk of lung cancer (OR = 1.41, 95% CI 1.14, 1.74) and asthma (OR = 1.05, 95% CI 0.98, 1.12). By contrast, each 1-SD unit higher BMI was associated with a reduced risk of FEV1 (β = − 0.09, 95% CI − 0.12, − 0.06) and FVC (β = − 0.17, 95% CI − 0.20, − 0.14), and a 1-SD lower genetically predicted cigarettes per day was associated with a decreased risk of FEV1 (β = − 0.08, 95% CI − 0.12, − 0.04), FVC (β = − 0.07, 95% CI − 0.11, − 0.02) and FEV1/FVC (β = − 0.04, 95% CI − 0.08, − 0.004).

Mediating effect of mediators in the association between education attainment and lung function and respiratory diseases

In the MVMR analysis of the impact of EA to lung function through the consumption of cigarettes per day, the direct effect of EA on FEV1, FVC and FEV1/FVC was β = 0.08 (95% CI 0.04, 0.13), 0.09 (95% CI 0.05, 0.14) and β = − 0.01 (95% CI − 0.05, 0.03) after adjusting for the number of cigarettes smoked per day (Fig. 3). The direct effect of BMI on FEV1, FVC and FEV1/FVC was − 0.09 (95% CI − 0.12, − 0.06), − 0.17 (95% CI − 0.02, − 0.14) and 0.11 (95% CI 0.09, 0.14), respectively, after accounting for EA. The proportion mediated of FEV1, FVC and FEV1/FVC by BMI was 15%, 23% and 379%, respectively (Table 3).

The MVMR analysis revealed that the direct effect of EA on lung cancer and asthma through cigarette consumption per day was 0.62 (95% CI 0.51, 0.76) and 0.85 (95% CI 0.78, 0.93) after adjusting for cigarettes smoked per day (Fig. 3). The direct effect of cigarettes per day on lung cancer and asthma was OR = 1.41 (95% CI 1.14, 1.74) and OR = 1.05 (95% CI 0.98, 1.12) after accounting for EA. The proportion mediated of lung cancer and asthma by cigarettes per day was 18% and 10% (Table 3).

Both smoking and BMI were included in the FEV1 outcome MVMR model when considered simultaneously, effect sizes for EA (β = 0.10, 95% CI 0.04, 0.16),

BMI (β = − 0.08, 95% CI − 0.11, − 0.05) and cigarettes per day (β = − 0.04, 95%CI − 0.09, 8e−06) (Fig. 3). Combined BMI and smoking mediated 44% of the effect of EA on FVC (Table 3). When BMI was the mediator, the effects of education on lung function and lung disease were shown in Fig. 3 and Table 3.

MR sensitivity analyses

According to the Cochran's Q test, the instrumental variables from education attainment to lung cancer did not show any heterogeneity, but there was heterogeneity in the other instrumental variables of the analysis which demonstrated a trend for the other instrumental variables (Table 4). In order to assess whether SNP has a horizontal pleiotropy, MR-Egger regression was used, which provided a valuable assessment of whether there was horizontal pleiotropy (Fig. 4). In the sensitivity analysis results, there was no significant evidence of directional pleiotropy (*P* > 0.05, Table 5). Furthermore, a further consistency between MR-weighted median and MR-IVW is in the direction of the distribution (Additional file 1: Table S1, Table 2). In reverse MR analyses between mediators and education attainment, the significant correlation between BMI and education attainment was found, but this reverse association could be due to horizontal pleiotropy (Egger intercept = − 0.0018; *P* = 0.0003). In terms of education attainment, Physical Activity and Cigarettes per day did not appear to have a causal effect (Additional file 1: Table S2). Moreover, leave-one-out analysis revealed that no SNP drove the results, and funnel plots were symmetrical (Fig. 4), indicating that the causal relationship has not been violated (Fig. 4). All SNPs have F-statistic ranging from 29.69 to 240.25. *F*-statistics > 10 considered suggestive of adequate instrument strength (Detailed information about SNPs is shown in Additional file 2: Table S3).
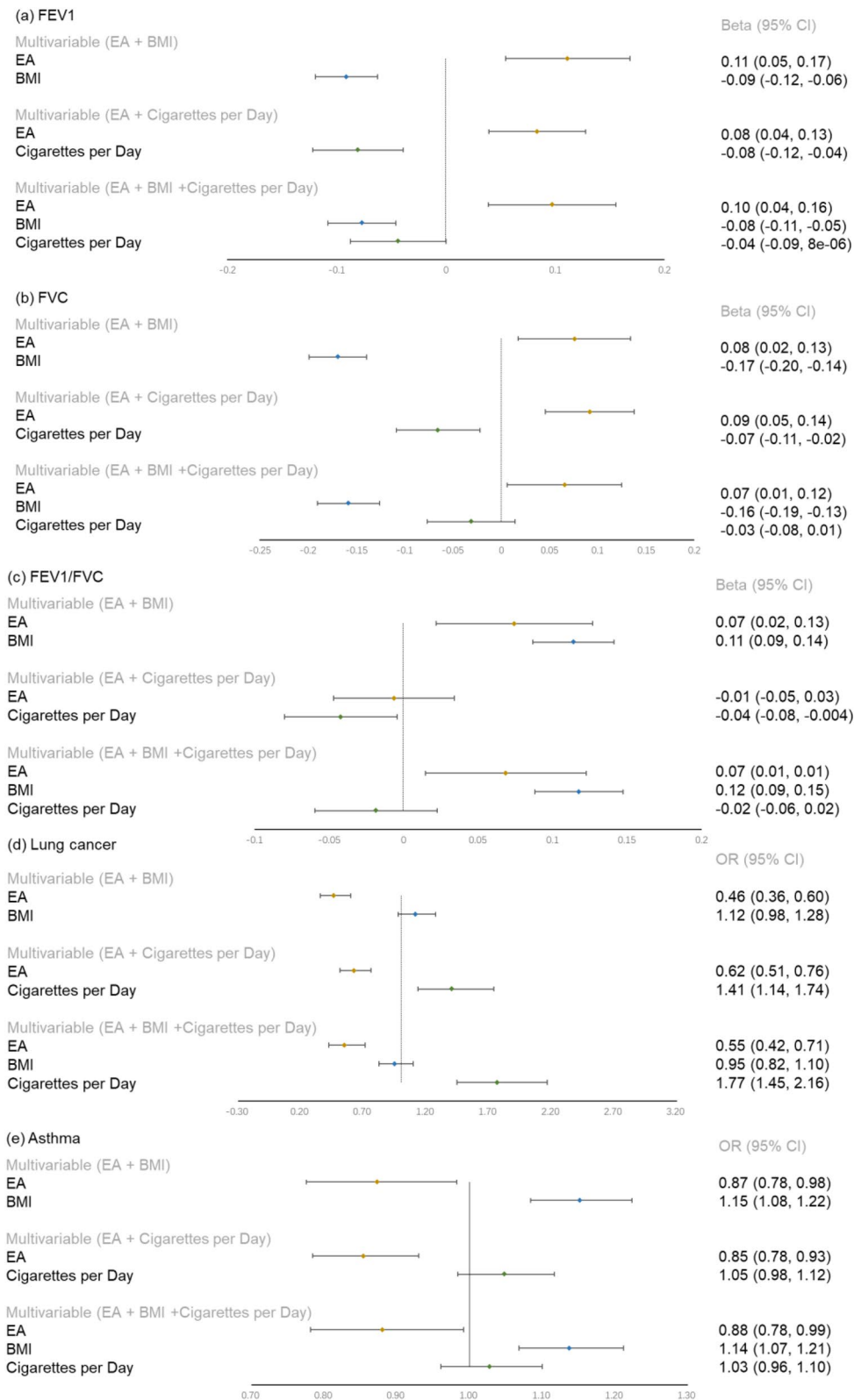
Lan *et al. Respiratory Research*   (2024) 25:115

Page 7 of 11

Fig. 3 Effect of one standard deviation (SD) increase in exposure on outcome in multivariable models. EA, educational attainment; *BMI* body mass index, *FEV1* forced expiratory volume in one second, *FVC* forced vital capacity, *FEV1/FVC* forced expiratory volume in one second/forced vital capacity

Lan *et al. Respiratory Research*   (2024) 25:115

Page 8 of 11

Table 3 Estimates of the effect of educational attainment on outcomes explained by each mediator and by both combined

Mediating pathway	Indirect effect (a × b)	Proportion mediated (%)
Educational attainment to FEV1		
Via BMI	0.02 (0.01, 0.02)	0.15
Via cigarettes per day	0.03 (0.01, 0.04)	0.26
Via BMI + cigarettes per day	0.003	0.03
Educational attainment to FVC		
Via BMI	0.03 (0.02, 0.04)	0.23
Via cigarettes per day	0.02 (0.01, 0.04)	0.18
Via BMI + cigarettes per day	0.05	0.44
Educational attainment to FEV1/ FVC		
Via BMI	− 0.02 (− 0.03, − 0.01)	3.79
Via cigarettes per day	0.01 (0.001, 0.03)	− 2.73
Via BMI + cigarettes per day	− 0.07	15.46
Educational attainment to lung cancer		
Via BMI	− 0.02 (− 0.04, 0.004)	0.03
Via cigarettes per day	− 0.11 (− 0.19, − 0.04)	0.18
Via BMI + cigarettes per day	− 0.02	0.03
Educational attainment to asthma		
Via BMI	− 0.02 (− 0.04, − 0.01)	0.15
Via cigarettes per day	− 0.02 (− 0.04, 0.005)	0.10
Via BMI + cigarettes per day	− 0.03	0.17

BMI body mass index, *FEV1* forced expiratory volume in one second, *FVC* forced vital capacity, *FEV1/FVC* forced expiratory volume in one second / forced vital capacity

Table 4 MR heterogeneity test of the association of educational attainment with each outcome and mediator

	Q statistic	Q-*p*-value
Educational attainment to FEV1		
MR heterogeneity test (MR-Egger)	824.0855	1.15e−61
MR heterogeneity test (IVW)	824.3595	1.89e−61
Educational attainment to FVC		
MR heterogeneity test (MR-Egger)	916.0814	4.70e−75
MR heterogeneity test (IVW)	916.8741	6.69e−75
Educational attainment to FEV1/FVC		
MR heterogeneity test (MR-Egger)	721.9336	1.93e−46
MR heterogeneity test (IVW)	725.8082	9.23e−47
Educational attainment to lung cancer		
MR heterogeneity test (MR-Egger)	259.45	0.4454708
MR heterogeneity test (IVW)	260.4956	0.4447875
Educational attainment to asthma		
MR heterogeneity test (MR-Egger)	461.4252	2.63e−13
MR heterogeneity test (IVW)	463.3352	2.30e−13
Educational attainment to BMI		
MR heterogeneity test (MR-Egger)	502.4243	1.62e−53
MR heterogeneity test (IVW)	502.5996	3.34e−53
Educational attainment to physical activity		
MR heterogeneity test (MR-Egger)	371.1704	8.48e−06
MR heterogeneity test (IVW)	371.9444	9.04e−06
Educational attainment to cigarettes per day		
MR heterogeneity test (MR-Egger)	465.8943	1.25e−16
MR heterogeneity test (IVW)	467.2486	1.25e−16

BMI body mass index, *FEV1* forced expiratory volume in one second, *FVC* forced vital capacity, *FEV1/FVC* forced expiratory volume in one second/forced vital capacity, *IVW* inverse variance weighted, *MR* Mendelian randomization

Discussion

In this MR study, the casual relationship between ET and respiratory functions and diseases was identified. To delve deeper into the mechanisms behind this association, we have identified three potential mediators from a pool of 23 modifiable risk factors. Our study findings reveal that education plays a crucial role in safeguarding lung function, preventing lung cancer, and mitigating the risk of asthma. An additional 4.2 years of schooling was associated with higher FEV1 and FVC values and lower lung cancer and asthma rates.

This is the first time that two-step MR analysis has been used to study the mediating relationship between EA and respiratory disease. Higher educational attainment is protective against respiratory disease, consistent with traditional observational findings. Actually, previous studies have shown that higher educational attainment has a protective effect on a range of health outcomes including lung cancer, artery stroke, type 2 diabetes. It is worth noting that this protective effect decreases as smoking and BMI are adjusted [28]. For example, smoking mediated 28% of the causal relationship between education and myocardial infarction, and BMI mediated 18%

[29]. This shows that the implementation of public health measures to reduce smoking and obesity has wide-ranging benefits in preventing the occurrence of disease.

In this study, although the protective effect of education on respiratory diseases was verified, the mediating factor of choice explained only one quarter of the effect of education, leaving a significant portion unexplained. There are a number of other factors that may explain the remaining associations, including poverty, employment, diet, psychosocial factors, and access to healthcare [30–34]. However, since many of these factors are not heritable and cannot be obtained in GWAS, they cannot be included in this study.

In the UK, it has been proven that raising the age at which students leave school can have an impact on EA and lead to improvements in population health and a decrease in mortality rates[35]. Although EA has often been used as a proxy for socioeconomic status in previous studies, it is important to acknowledge that interventions solely targeting educational attainment may not offer an optimal solution for alleviating the burden of
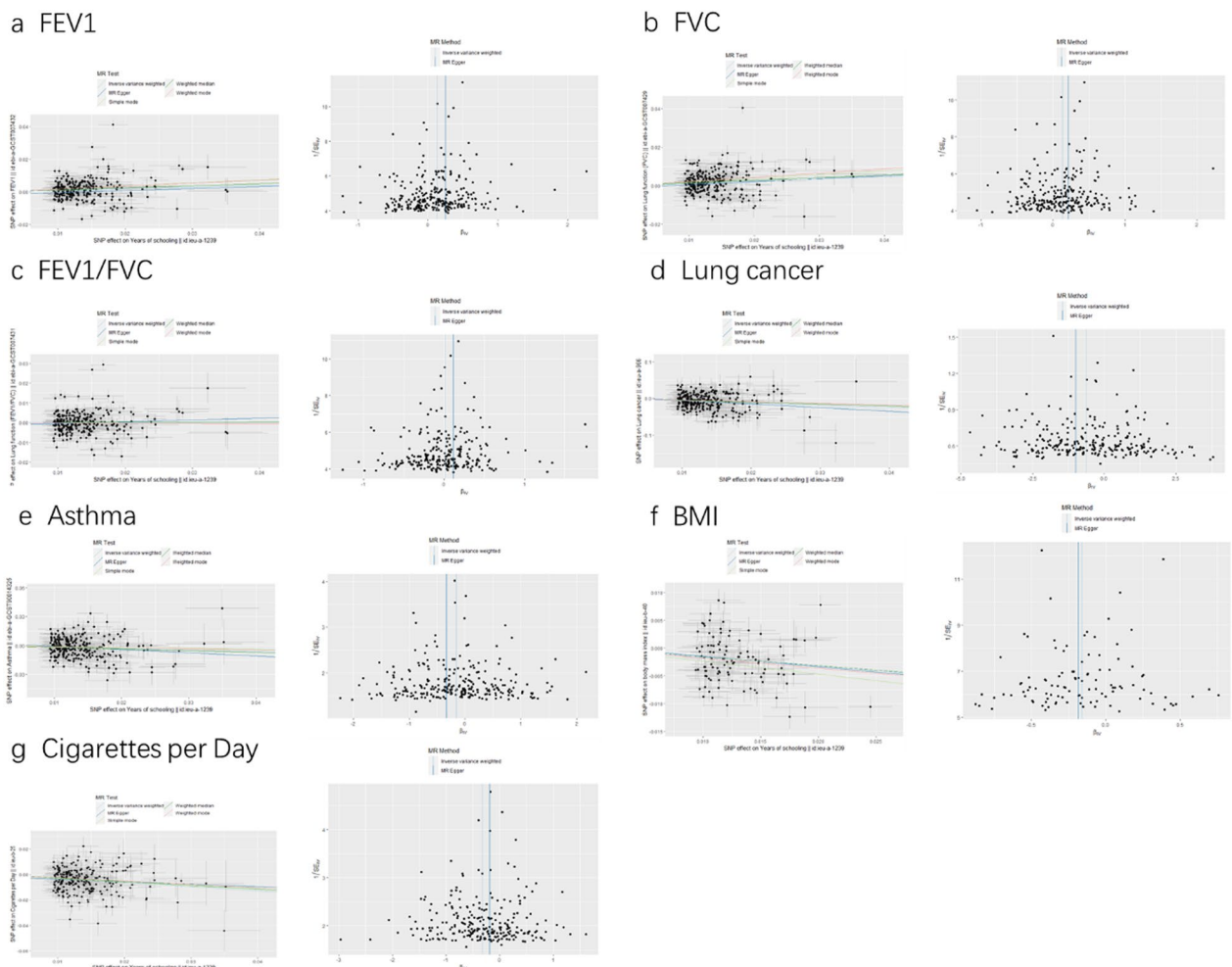
Fig. 4 Mendelian randomization scatterplots and funnel plots of educational attainment to each mediator and outcome association. *BMI* body mass index, *FEV1* forced expiratory volume in one second, *FVC* forced vital capacity, *FEV1/FVC* forced expiratory volume in one second/forced vital capacity

Table 5 MR directional pleiotropy test (MR Egger) of the association of educational attainment with each outcome and mediator

Exposure	Outcome	Egger intercept	SE	*P* value
EA	FVC	− 0.0006	0.0012	0.6383
	FEV1	− 0.0013	0.0011	0.2441
	FEV1/FVC	− 0.0004	0.0012	0.2750
	Lung cancer	0.0052	0.0051	0.3098
	Asthma	0.0025	0.0024	0.300
	BMI	0.0004	0.0019	0.8493
	Physical activity	0.0007	0.0010	0.4613
	Cigarettes per day	− 0.0018	0.0022	0.4044

BMI body mass index, *FEV1* forced expiratory volume in one second, *FVC* forced vital capacity, *FEV1/FVC* forced expiratory volume in one second/forced vital capacity, *SE* standard error

respiratory disease. In this study, a two-stage MR study was conducted to demonstrate that some risk factors mediate the relationship between EA and respiratory disease, and that these factors are more likely to change than EA.

In comparison to prior investigations, this study encompasses the following commendable attributes: (1) The study uses SNPs as genetic instruments can capture the impact of genetic variation on the phenotype or disease of interest. This approach effectively mitigates the confounding effects of reverse causality and errors. Due to allele random assignment at the time of conception, MR results that are insensitive to reverse causation. Additionally, using SNP as a tool variable can also improve the reliability and accuracy of MR analysis. (2) Exposure and outcome summary statistics in the study were obtained from the largest and most recent GWAS.

Lan *et al. Respiratory Research*   (2024) 25:115

Page 10 of 11

(3) In order to improve the statistical power, a rigorous screening process was carried out for IVs (4) As part of the research process, multiple sensitivity analyses were performed in order to improve the results' accuracy. Furthermore, the MR analysis results align with those of observational studies, thereby reinforcing the robustness of the conclusions.

Notwithstanding the aforementioned strengths, this study is subject to some limitations that warrant consideration. Firstly, the GWAS used in the study exclusively featured on European populations. Thus, the generalization of results is not suitable for non-European people. Therefore, newer GWAS studies should focus on non-European populations. Secondly, given that EA has sex differences with respiratory diseases, associations and mediations may also differ between the sexes. However, as GWAS summary data were used, the effects of sex and age on outcomes could not be studied. A sex-stratified GWAS study may be used in future MR studies to address this issue. Thirdly, since lung cancer and asthma are binary variables, log-odds should be used in MR Analyses. The optimality of this approach is not achieved since the odds ratios do not collapse, i.e. marginal ORs are not equivalent to conditional odds ratios. Fourthly, the GWAS summary data used in this article comes from different repositories, in which case there is some heterogeneity between the data. This is inevitable, however, because when different data sources are selected, the bias of instrumental variables is reduced and the reliability of the results is improved. Finally, there is a possibility that GWAS results may be biased by sample overlap between studies.

Conclusion

Elevated levels of EA may potentially exert a protective effect on respiratory diseases, with modifiable risk factors such as BMI and cigarettes per day mediating this relationship. Interventions to reduce smoking and adiposity may reduce much of this risk, which assumes even greater significance for individuals with respiratory disease. However, most of the remaining effects of EA on the relationship between respiratory disease remain unexplained. As such, there is a pressing need for enhanced preventive measures to address socioeconomic and educational disparities, as well as further research into other modifiable risk factors.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12931-024-02722-4.

Additional file 1: Figure S1. Overview of the process of identifying the mediators. **Table S1.** Mendelian randomization analysis of the effect of

educational attainment on lung function and disease. **Table S2.** Reverse MR analysis of mediators to education attainment.

Additional file 2: Table S3. All instrumental variables used in Mendelian randomization analysis.

Declarations

Ethics approval and consent to participate
Not applicable.

Competing interests
All the authors declare no conflicting interests.

Author details
[1]Department of Epidemiology and Health Statistics, School of Public Health, Fujian Medical University, Fuzhou, China. [2]The Second Clinical Medical School, Nanchang University, Nanchang, China. [3]Clinical Research Unit, The Second Affiliated Hospital, Fujian Medical University, Quanzhou, China. [4]The First Clinical Medical School, Shanxi Medical University, Taiyuan, China. [5]Department of Pulmonary and Critical Care Medicine, Center of Respiratory Medicine of Fujian Province, The Second Affiliated Hospital of Fujian Medical University, Quanzhou, China.

References

1. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet. 2020;396(10258):1204–1222. https://doi.org/10.1016/S0140-6736(20)30925-9 (Erratum in: Lancet. 2020;396(10262):1562).
2. Mannino DM, et al. Lung function and mortality in the United States: data from the First National Health and Nutrition Examination Survey follow up study. Thorax. 2003;58(5):388–93.
3. Blane D, et al. Association of cardiovascular disease risk factors with socioeconomic position during childhood and during adulthood. BMJ. 1996;313(7070):1434–8.
4. Davey Smith G, et al. Education and occupational social class: which is the more important indicator of mortality risk? J Epidemiol Community Health. 1998;52(3):153–60.

5.   Marang-vandeMheen PJ, et al. Socioeconomic differentials in mortality among men within Great Britain: time trends and contributory causes. J Epidemiol Community Health. 1998;52(4):214–8.

6.   Martikainen P, et al. Educational differences in lung cancer mortality in male smokers. Int J Epidemiol. 2001;30(2):264–7.

7.   Smith GD, et al. Lifetime socioeconomic position and mortality: prospective observational study. BMJ. 1997;314(7080):547–52.

8.   Cohen AK, Syme SL. Education: a missed opportunity for public health intervention. Am J Public Health. 2013;103(6):997–1001.

9.   Zhang DD, et al. Association between socioeconomic status and chronic obstructive pulmonary disease in Jiangsu province, China: a population-based study. Chin Med J (Engl). 2021;134(13):1552–60.

10. Swaminathan R, et al. Education and cancer incidence in a rural population in south India. Cancer Epidemiol. 2009;33(2):89–93.

11. Smith GD, Ebrahim S. "Mendelian randomization": can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol. 2003;32(1):1–22.

12. Sanderson E. Multivariable mendelian randomization and mediation. Cold Spring Harb Perspect Med. 2021;11(2):a038984. https://doi.org/10.1101/cshperspect.a038984.

13. Li Y, et al. Evaluating the causal association between educational attainment and asthma using a Mendelian randomization design. Front Genet. 2021;12: 716364.

14. Lee JJ, 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.

15. Shrine N, et al. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. Nat Genet. 2019;51(3):481–93.

16. Wang Y, et al. Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. Nat Genet. 2014;46(7):736–41.

17. Valette K, et al. Prioritization of candidate causal genes for asthma in susceptibility loci derived from UK Biobank. Commun Biol. 2021;4(1):700.

18. Yengo L, et al. Meta-analysis of genome-wide association studies for height and body mass index in ∼700000 individuals of European ancestry. Hum Mol Genet. 2018;27(20):3641–9.

19. Tyrrell J, et al. Genetic predictors of participation in optional components of UK Biobank. Nat Commun. 2021;12(1):886.

20. Liu M, et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. Nat Genet. 2019;51(2):237–44.

21. Hemani G, Bowden J, Davey Smith G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. Hum Mol Genet. 2018;27(2):195–208.

22. Lawlor DA, et al. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med. 2008;27(8):1133–63.

23. Burgess S, Thompson SG. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. Am J Epidemiol. 2015;181(4):251–60.

24. Carter AR, et al. Mendelian randomisation for mediation analysis: current methods and challenges for implementation. Eur J Epidemiol. 2021;36(5):465–78.

25. VanderWeele TJ. Mediation analysis: a practitioner's guide. Annu Rev Public Health. 2016;37:17–32.

26. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, Laurin C, Burgess S, Bowden J, Langdon R, Tan VY, Yarmolinsky J, Shihab HA, Timpson NJ, Evans DM, Relton C, Martin RM, Davey Smith G, Gaunt TR, Haycock PC. The MR-Base platform supports systematic causal inference across the human phenome. Elife. 2018;7:e34408. https://doi.org/10.7554/eLife.34408

27. Dean CB, Nielsen JD. Generalized linear mixed models: a review and some extensions. Lifetime Data Anal. 2007;13(4):497–512.

28. 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.

29. Carter AR, et al. Understanding the consequences of education inequality on cardiovascular disease: mendelian randomisation study. BMJ. 2019;365:l855.

30. Belot A, et al. Association between age, deprivation and specific comorbid conditions and the receipt of major surgery in patients with non-small cell lung cancer in England: a population-based study. Thorax. 2019;74(1):51–9.

31. Brundisini F, et al. Chronic disease patients' experiences with accessing health care in rural and remote areas: a systematic review and qualitative meta-synthesis. Ont Health Technol Assess Ser. 2013;13(15):1–33.

32. Collins PF, et al. Influence of deprivation on health care use, health care costs, and mortality in COPD. Int J Chron Obstruct Pulmon Dis. 2018;13:1289–96.

33. Scoditti E, Massaro M, Garbarino S, Toraldo DM. Role of diet in chronic obstructive pulmonary disease prevention and treatment. Nutrients. 2019;11(6):1357. https://doi.org/10.3390/nu11061357.

34. Watson A, Wilkinson TMA. Digital healthcare in COPD management: a narrative review on the advantages, pitfalls, and need for further research. Ther Adv Respir Dis. 2022;16:17534666221075492.

35. Davies NM, et al. The causal effects of education on health outcomes in the UK Biobank. Nat Hum Behav. 2018;2(2):117–25.

Publisher's Note

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