Original Article

Check for updates

Lipids, Anthropometric Measures, Smoking and Physical Activity Mediate the Causal Pathway From Education to Breast Cancer in Women: A Mendelian Randomization Study

Hongkai Li [1,2,*], Lei Hou [1,2,*], Yuanyuan Yu [1,2], Xiaoru Sun [1,2], Xinhui Liu [1,2], Yifan Yu [1,2], Sijia Wu [1,2], Yina He [1,2], Yutong Wu [1,2], Li He [1,2], Fuzhong Xue [1,2]

[1]Department of Epidemiology and Health Statistics, School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan, People's Republic of China
[2]Institute for Medical Dataology, Cheeloo College of Medicine, Shandong University, Jinan, People's Republic of China

Correspondence to
Fuzhong Xue

Department of Epidemiology and Health Statistics, School of Public Health, Cheeloo College of Medicine, Shandong University, Shandong University, 44 Wenhua West Road, Jinan, Shandong 250012, China.
E-mail: xuefzh@sdu.edu.cn

*Hongkai Li and Lei Hou are co-first authors of this article and have contributed equally to this work.

ORCID iDs
Hongkai Li
https://orcid.org/0000-0003-1848-937X
Lei Hou
https://orcid.org/0000-0002-5636-4028
Yuanyuan Yu
https://orcid.org/0000-0001-6304-0817
Xiaoru Sun
https://orcid.org/0000-0003-4435-4348
Xinhui Liu
https://orcid.org/0000-0002-7894-1172
Yifan Yu
https://orcid.org/0000-0002-9741-0051
Sijia Wu
https://orcid.org/0000-0002-9767-9966

ABSTRACT

Purpose: We aimed to investigate whether obtaining a higher level of education was causally associated with lower breast cancer risk and to identify the causal mechanism linking them.
Methods: The main data analysis used publicly available summary-level data from 2 large genome-wide association study consortia. Mendelian randomization (MR) analysis used 65 genetic variants derived from the Social Science Genetic Association Consortium as instrumental variables for years of schooling. The outcomes from the Breast Cancer Association Consortium (BCAC) were the overall breast cancer risk (122,977 cases/105,974 controls in women) and the two subtypes: estrogen receptor (ER)-positive breast cancer and ER-negative breast cancer. Fixed and random effects inverse variance weighted methods were used to estimate the causal effects, along with other additional MR methods for sensitivity analyses.
Results: Results showed that each additional standard deviation of 4.2 years of education was causally associated with a 27% lower risk of ER-negative breast cancer (odds ratio, 0.73; 95% confidence interval, 0.64–0.84; p-value < 0.001). This finding was consistent with the results of the sensitivity analyses. Physical activities can help improve the protective effect of education against breast cancer, with relatively large mediation proportions. Education increases the risk of ER-positive breast cancer due to alterations in high-density lipoprotein level, triglyceride level, height, waist-to-hip ratio, body mass index, and smoking status, with relative medium mediation proportions. Other mediators including low-density lipoprotein, hip circumference, number of cigarettes smoked per day, time spent performing light physical activity, and performing vigorous physical activity for > 10 minutes explain a small part of the causal effect of education on the risk of developing breast cancer, and their mediation proportion is approximately 1%.
Conclusion: A low level of education is a causal risk factor in the development of breast cancer as it is associated with poor lipid profile, obesity, smoking, and types of physical activity.

Keywords: Breast Neoplasms; Education; Mediation Analysis; Mendelian Randomization Analysis; Meta-Analysis

Yina He (ID)
https://orcid.org/0000-0002-8929-3692
Yutong Wu (ID)
https://orcid.org/0000-0002-4258-2032
Li He (ID)
https://orcid.org/0000-0003-4028-3191
Fuzhong Xue (ID)
https://orcid.org/0000-0003-0378-7956

Conflict of Interest
The authors declare that they have no competing interests.

Author Contributions
Conceptualization: Li H, Xue F; Data curation: Wu S, He Y, Wu Y, He L; Formal analysis: Li H, Hou L; Investigation: Yu Y; Methodology: Hou L; Project administration: Li H; Resources: Sun X; Supervision: Yu Y; Visualization: Liu X; Writing - original draft: Hou L

INTRODUCTION

Breast cancer is the second leading cause of mortality among European women, and almost one in eight women develop breast cancer during their lifetime [1,2]. Each year, approximately 17 million new cases have been reported [3,4]. With the increasing burden of breast cancer, it is imperative to identify the modifiable risk factors for prevention. Education is a key component of socioeconomic status and may lower the breast cancer risk by altering the lipid profile, anthropometric measurements, physical activity, smoking, etc. [5]. Numerous observational studies have investigated the relationship between education and breast cancer, but the results have been inconsistent [6-10]. For instance, a case-control study and a cohort study suggested opposite results regarding the relationship between education level and breast cancer [11,12]. The former found an inverse association between educational level and breast cancer risk (odds ratio [OR], 0.17), whereas the latter showed that, in contrast to women who completed less than 9 years of education, university graduates had a higher probability of being diagnosed with in situ (hazard ratio [HR], 1.44) and invasive breast cancer (HR, 1.28). These contradictory results may be attributed to the limitations of traditional observational studies, including unmeasured confounding factors and reverse causation.

The instrumental variable (IV) method exploits a natural experiment to determine the causal association between an exposure and an outcome. A valid instrument must satisfy the following 3 assumptions: 1) Relevance: IV (G) is robustly related to exposure (X); 2) Exchangeability: IV (G) is independent of any unobserved confounders (U) of the exposure and outcome relationship; and 3) Exclusion restriction: IV (G) affects the outcome (Y) only through the exposure (X) [12]. In observational data, the use of genetic variants as instrumental variables has been termed as "Mendelian randomization (MR)." MR analyses using summarized data have recently become popular because of the large number of published genome-wide association studies (GWAS) on large sample populations that are publicly available [13], significantly increasing its statistical power.

Using the MR approach, several studies have found that the level of education was causally associated with myopia, lung cancer, and coronary heart disease [13-15]. However, to the best of our knowledge, this is the first study to report the causal relationship between educational attainment and breast cancer. In this study, we used 2-sample MR analyses to identify the potential causation between education level and breast cancer and its estrogen receptor (ER) subtypes. Furthermore, we investigated the causal pathways that link between them.

METHODS

Genetic variants related to educational attainment

Educational attainment (EA) was measured as the number of years of schooling completed. A large genetic association study reported by Lee et al. [16] identified 317 single nucleotide polymorphisms (SNPs) robustly associated with educational attainment in the Social Science Genetic Association Consortium (SSGAC) at a GWAS threshold of statistical significance (of 766 participants, 345 were of European descent; p-value $< 5 \times 10^{-8}$; linkage disequilibrium [LD] $r^2 < 0.001$) (**Table 1**). These 317 SNPs explain 2.03% of the variations in educational attainment across individuals. The F statistic was larger than the "rule of thumb" of 10 [12], which means that the instruments used strongly predict the educational attainment. Thus, it

Table 1. Details of the studies included in the Mendelian randomization analyses

Consortium	Phenotype	Participant	Year	Web source
SSGAC	Years of schooling	766,345	2018	https://www.thessgac.org/
BCAC	Breast cancer	228,951	2017	https://bcac.ccge.medschl.cam.ac.uk/
GLGC	Lipids	188,557	2013	http://csg.sph.umich.edu/willer/public/lipids2013/
GIANT	Anthropometric measures	224,459	2015	http://portals.broadinstitute.org/collaboration/giant
UK Biobank	BMI	461,460	2018	https://www.ukbiobank.ac.uk/
UK Biobank	Sleep duration	128,266	2018	https://www.ukbiobank.ac.uk/
MRC-IEU	Physical activities	160,376	2018	https://gwas.mrcieu.ac.uk/
TAG	Smoke	74,053	2010	https://www.med.unc.edu/pgc/download-results/

SSGAC = Social Science Genetic Association Consortium; BCAC = Breast Cancer Association Consortium; GLGC = Global Lipids Genetics Consortium; GIANT = Genetic Investigation of ANthropometric Traits; MRC-IEU = MRC Integrative Epidemiology Unit; TAG = Tobacco and Genetics consortium.

is sufficient to generate a strong genetic instrument based on these 317 SNPs. In this study, we only used 317 SNPs and summarized the data collected from the SSGAC.

GWAS summary level data on breast cancer

The GWAS summary data of breast cancer individuals of European descent were retrieved from the BCAC database (**Table 1**) [17]. Results were available for 291 of the 317 EA-associated leading SNPs for the following breast cancer subtypes (**Supplementary Table 1**): overall breast cancer (122,977 cases/105,974 controls), ER-positive breast cancer (69,501 cases/105,974 controls), and ER-negative breast cancer (21,468 cases/105,974 controls). Ten palindromic SNPs with intermediate allele frequencies (rs12134151, rs13130765, rs1455350, rs2414072, rs2478208, rs2545798, rs320693, rs60483752, rs6867851, and rs7920624) were removed from the analysis. We used the summary data from the following 4 databases: 1) OncoArray Consortium (61,282 cases and 45,494 controls), 2) Collaborative Oncological Gene-environment Study (iCOGS: 46,785 cases and 42,892 controls), 3) 11 other breast cancer genome-wide association studies (GWAS; 14,910 cases and 17,588 controls), and 4) a combination of the above three databases.

Other breast cancer risk factor data

Summary results from genome-wide association meta-analyses for lipids were obtained from 4 genetic consortia, including high-density lipoprotein (HDL) and low-density lipoprotein (LDL) and total cholesterol, triglycerides (TGs), anthropometric measurements (hip and waist circumference, waist-to-hip ratio [WHR], height measurement, and body mass index [BMI]), smoking, sleep duration, and physical activity [18]. The websites used for data collection and consortia are listed in **Table 1**. Only the summary statistics of patients of European descent were obtained from the analyses. The statistical analyses included linear/logistic regression coefficients (beta/log [OR]), standard errors, and *p*-values for the genetic association analysis.

The main steps of the study are presented in **Figure 1**. Multiple MR approaches have been used to obtain the estimates of educational attainment for breast cancer and its ER subtypes. We conducted a fixed and random effects inverse variance weighted (IVW) meta-analysis [12,19] of the Wald ratio for individual SNPs. Heterogeneity was detected in the Wald ratio; if heterogeneity exists, the random effects IVW is a better method; however, we used fixed IVW. The IVW method assumes that all SNPs are valid instruments that satisfy the three core assumptions in MR. Three additional MR methods were also used as sensitivity analyses to assess the robustness of the results: MR-Egger regression, weighted median, and weighted mode methods. The intercept of the MR-Egger regression provides an estimate of the average pleiotropic effect of all SNPs. If it differs from zero, it indicates the presence of directional
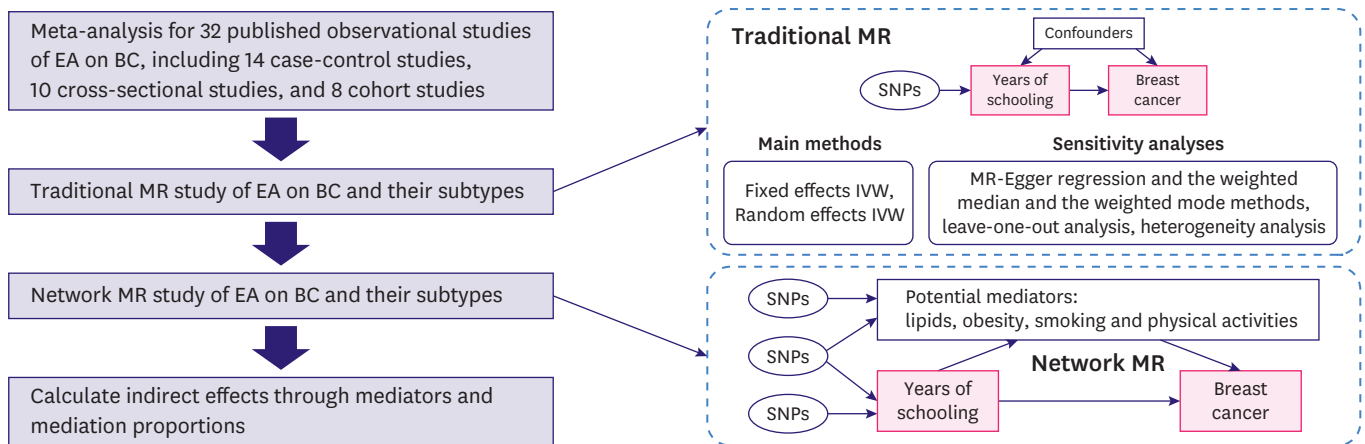
Figure 1. Research design.
EA = educational attainment; BC = breast cancer; MR = Mendelian randomization; SNP = single nucleotide polymorphism; IVW = inverse variance weighting.

pleiotropy. An SNP with directional pleiotropy implies that there is an alternative causal pathway from the genetic variant to the outcome, except for that via the risk factor. In this case, the third assumption in the MR (exclusion restriction) is violated. We also performed a leave-one-out analysis in which we sequentially omitted one SNP at a time to determine whether the MR estimate was driven or biased by a single SNP.

To investigate the potential mechanisms involved in the association between education and breast cancer, we applied a network MR to explore the potential mediators of this causal pathway. We selected 25 potential mediators based on the existing literature, including lipids (HDL, LDL, total cholesterol, and TGs), anthropometric measures (waist and hip circumference, WHR, height measurement, and BMI), smoking, sleep duration, and physical activities, which can be risk factors for breast cancer. MR was initially performed to estimate the causal effects of educational attainment on these risk factors. Additional MR analyses were performed to determine the risk factors of breast cancer if EA showed a causal effect on the above risk factors. Finally, we calculated the indirect effects of each mediator and their mediation proportions (MPs). Details of statistical methods are illustrated in **Supplementary Data 1**.

Based on a simulation study [20] on sample overlap and the degree of bias in the MR analysis, a less than 5% degree of overlap was not considered significant. The proportion of sample overlap in the summary data used in the MR analyses was within the acceptable range and did not lead to an estimation bias. The R package *TwoSampleMR* (v0.5.1) was used to perform all of the above MR analyses (version 3.6.3). The calculation of power can be found at http://cnsgenomics.com/shiny/mRnd/.

RESULTS

Meta-analysis of the impact of educational attainment on the risk of breast cancer

First, we performed a meta-analysis of all published observational studies that explored the relationship between educational attainment levels and breast cancer. We searched PubMed, MEDLINE, Embase, and Web of Science for studies that used the term "education" or "schooling" and "breast cancer" from inception to October 21, 2020. We excluded

publications that 1) were conference abstracts, letters, commentaries, editorials, reviews, study proposals, or theoretical papers; 2) whose primary exposure variable was not education; and 3) set education as an outcome. After applying our inclusion and exclusion criteria, 32 MR studies (**Supplementary Table 1**) were included in the meta-analysis.

We then pooled the study-specific estimates using a random-effects model for the meta-analysis. The forest plots of the meta-analysis are shown in **Supplementary Figures 1-3**. The articles included 14 case-control studies, 10 cross-sectional studies, and 8 cohort studies. We evaluated the study heterogeneity by calculating the I^2 statistic using Cochran's Q test. Significant heterogeneity between these studies was found after calculating the I^2 value using Cochran's Q test. Then, a random-effects model was performed for the meta-analysis. The meta-analysis results from the 3 studies revealed a positive association between educational attainment and breast cancer.

Causal effect of educational attainment on the risk of developing breast cancer

A large heterogeneity was found in several databases; thus, a random effects IVW was performed. For databases with no heterogeneity, a fixed-effects IVW was used. The results of the heterogeneity tests are listed in **Table 2**. No directional pleiotropy was found in any of the analyses performed.

In the combined dataset, genetically predicted higher educational attainment tended to decrease the risk of ER-negative breast cancer (**Figure 2**). Using the random-effects IVW method, each additional standard deviation (SD) higher education was associated with a 27% lower risk of ER-negative breast cancer (OR, 0.73; 95% confidence interval [CI], 0.64–0.84; $p < 0.001$). **Supplementary Figure 4** shows the forest plot of 291 SNPs associated with educational attainment and the risk of ER-negative breast cancer. As expected, the associations were consistent with the results of the sensitivity analyses using the weighted mode (OR, 0.76; 95% CI, 0.64–0.91; $p = 0.002$) and MR-Egger method (OR, 0.49; 95% CI, 0.29–0.84; $p = 0.01$), but provided less precise estimates than the IVW method. Scatter plots are shown in **Supplementary Figure 5**. The heterogeneity test showed that no single SNP

Table 2. Heterogeneity test and MR-Egger pleiotropy test of the causal effects of educational attainment on the risk of developing breast cancer and its subtypes

Outcome	Heterogeneity test				Pleiotropy test	
	IVW		MR-Egger		MR-Egger	
	Q	p-value	Q	p-value	Intercept	p-value
Breast cancer						
Combination	622.660	< 0.001*	620.082	< 0.001*	0.003	0.274
OncoArray	409.585	< 0.001*	405.472	0.001*	0.005	0.088
iCOGS	423.556	< 0.001*	423.517	0.010*	−0.001	0.870
GWAS	376.679	< 0.001*	375.888	0.809	0.004	0.436
ER+						
Combination	525.272	< 0.001*	523.770	< 0.001*	0.002	0.363
OncoArray	380.157	< 0.001*	377.314	< 0.001*	0.005	0.141
iCOGS	402.683	< 0.001*	402.556	< 0.001*	0.001	0.763
GWAS	316.552	0.136	315.444	0.137	−0.009	0.314
ER−						
Combination	417.539	< 0.001*	414.293	< 0.001*	0.006	0.134
OncoArray	339.408	0.024*	337.396	0.026*	0.006	0.190
iCOGS	323.401	0.086	323.248	0.081	0.002	0.712
GWAS	326.926	0.067	325.029	0.071	0.010	0.195

MR = Mendelian randomization; IVW = inverse variance weighted; Breast cancer = overall breast cancer risk; ER+ = estrogen receptor-positive breast cancer risk; ER− = estrogen receptor-negative breast cancer risk; OncoArray = OncoArray Consortium; iCOGS = international Collaborative Oncological Gene-environment Study; GWAS = 11 other breast cancer genome-wide association studies; Combination = combination of above 3 databases.
*The p-values < 0.05 are statistically significant.
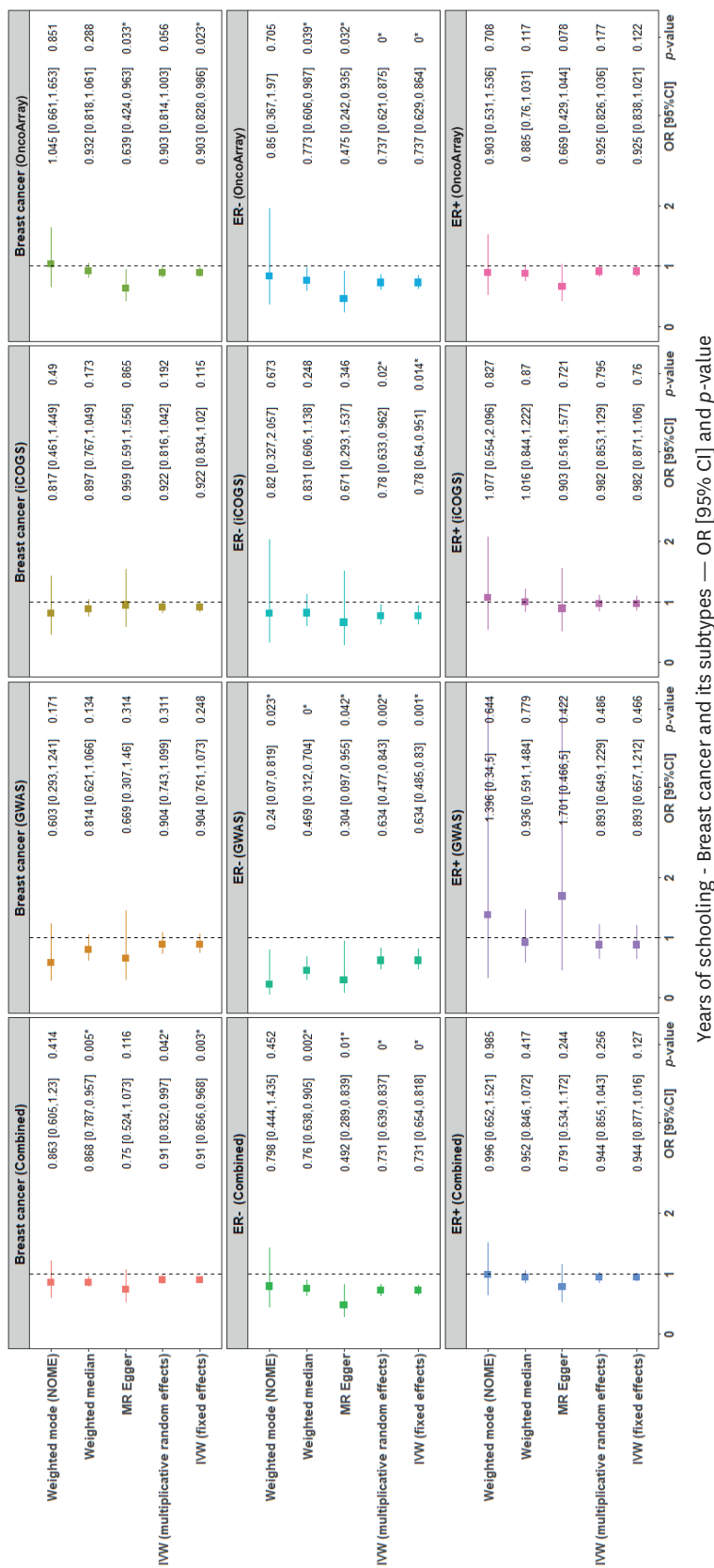
Figure 2. Causal effects of the level of education on the risk of breast cancer and estrogen receptor subtypes.

Breast cancer = overall breast cancer risk; ER⁺ = estrogen receptor-positive breast cancer risk; ER⁻ = estrogen receptor-negative breast cancer risk; OncoArray = OncoArray Consortium; iCOGS = international Collaborative Oncological Gene-environment Study; GWAS = 11 other breast cancer genome-wide association studies; Combination = combination of above 3 databases; MR = Mendelian randomization; IVW = inverse variance weighted; OR = odds ratio; CI = confidence interval.

significantly contributed to the overall effect of education on the risk of ER-negative breast cancer (**Supplementary Figure 6**). The results of the MR-Egger test suggested that there was no directional pleiotropy (**Table 2**) and were consistent with those of the IVW analysis. Similar results were obtained using the OncoArray, iCOGS, and GWAS datasets (**Figure 2** and **Supplementary Figures 7-15**). In addition, we found a very weak causal association for overall breast cancer (OR, 0.91; 95% CI, 0.83–0.997; p = 0.042) in the combined database but a null causal association in the other three datasets (**Figure 2**). However, we observed a null causal association for ER-positive breast cancer (OR, 0.94; 95% CI, 0.85–1.04; p = 0.26) (**Figure 2**) in all databases. The results of the pleiotropy tests are listed in **Table 2**.

Causal effects of education on the potential risk factors of breast cancer

To identify the underlying mechanism of the association between the level of education and ER-negative breast cancer, we investigated whether several potential cancer risk factors play a role. We found that education had causal effects on 20 out of the 24 risk factors. **Figure 3** shows that each SD higher level of education was associated with 32% lower odds of smoking, 1.89 times higher odds of smoking cessation among smokers, less smoking intensity (−2.26 [−3.48 to −0.65] cigarettes per day), 0.35 lower BMI, 0.13 lower WHR, 0.35 higher height, 0.09 higher hip circumference, 0.15 mmol/L lower TGs, and 0.16 mmol/L higher HDL-cholesterol (p <0.05).

In addition, physical activity performed during the last four weeks has a causal effect on the risk of ER-negative breast cancer. Every increase in the level of education was associated with 13% higher odds of performing light do-it-yourself (DIY) activities (e.g., pruning and watering the lawn) and 8% higher odds of performing heavy DIY activities (e.g., weeding, lawn mowing, carpentry, and digging). It was also associated with 6% higher odds of performing strenuous sports, 12% higher odds of performing leisure walking, and 16% higher odds of performing other activities (e.g., swimming, cycling, keeping fit, and bowling) in the last four weeks. In addition, the risk of ER-negative breast cancer decreased when performing > 10 minutes of moderate (0.24 days/week) and vigorous (0.07 days/week) physical activities.

The results of other sensitivity analyses for the above causal associations are shown in **Supplementary Figures 16** and **17**. All samples used provided sufficient statistical power (100%) to identify the causal effects.

Indirect effects of education on breast cancer through the mediators

MR was performed to evaluate the causal effects of the 20 potential mediators on the risk of breast cancer (**Supplementary Figures 18-21**). We calculated the indirect effects of education on ER-negative breast cancer through these mediators (**Table 3**). For continuous mediators, lipids, obesity, and physical activities play important roles, while the directions of the indirect effects through BMI (OR, 1.073; MP, 6.05%), WHR (OR, 1.021; MP, 1.743%), HDL (OR, 1.023; MP, 1.962%), TGs (OR, 1.022; MP, 2.544%), time spent engaging in vigorous physical activities (OR, 1.03; MP, 1.894%), and performance of moderate physical activities > 10 minutes (OR, 1.032; MP, 2.664%) were opposite to the total effect. Other mediators including LDL, hip circumference, number of cigarettes smoked per day, height measurement, engaging in vigorous physical activities for > 10 minutes, and time spent performing light physical activities explained a small part of the causal effect of education on the risk of developing ER-negative breast cancer, and their MP was less than 1%. By contrast, all binary mediators had large MPs. Performance of light DIY activities, walking for pleasure, and engaging in strenuous sports can help improve the protective effect of education against

Outcome		nsnp	b	95% CI	p	Q_i	Q_e	Egger
Cigarettes smoked per day		247	−2.257	[−3.134, −1.381]	0.000	0.407	0.390	0.948
Time spent doing moderate physical activity		307	−0.056	[−0.137, 0.025]	0.173	0.096	0.103	0.168
Time spent doing light physical activity		307	−2.000	[−2.098, −1.902]	0.001	0.475	0.488	0.184
Sleep duration		306	0.042	[−0.004, 0.089]	0.073	0.000	0.000	0.479
HDL-C		227	0.159	[0.096, 0.221]	0.000	0.000	0.000	0.761
LDL-C		227	−0.081	[−0.150, −0.012]	0.022	0.000	0.000	0.623
Total cholesterol		227	−0.051	[−0.123, 0.021]	0.162	0.000	0.000	0.776
TGs		227	−0.147	[−0.209, −0.084]	0.000	0.000	0.000	0.773
Hip circumference		247	0.088	[0.014, 0.161]	0.020	0.000	0.000	0.352
Waist circumference		247	−0.061	[−0.128, 0.005]	0.071	0.000	0.000	0.575
WHR		247	−0.126	[−0.187, −0.064]	0.000	0.005	0.005	0.902
Height		249	0.101	[0.019, 0.184]	0.016	0.000	0.000	0.504
Time spent doing vigorous physical activity		307	−0.231	[−0.310, −0.152]	0.000	0.011	0.011	0.402
Vigorous physical activity 10+ minutes		307	−0.067	[−0.127, −0.007]	0.028	0.000	0.000	0.149
BMI		307	−0.346	[−0.397, −0.295]	0.000	0.000	0.000	0.739
Moderate physical activity 10+ minutes		307	−0.237	[−0.311, −0.162]	0.000	0.000	0.000	0.267

AR [95% CI] (continious)

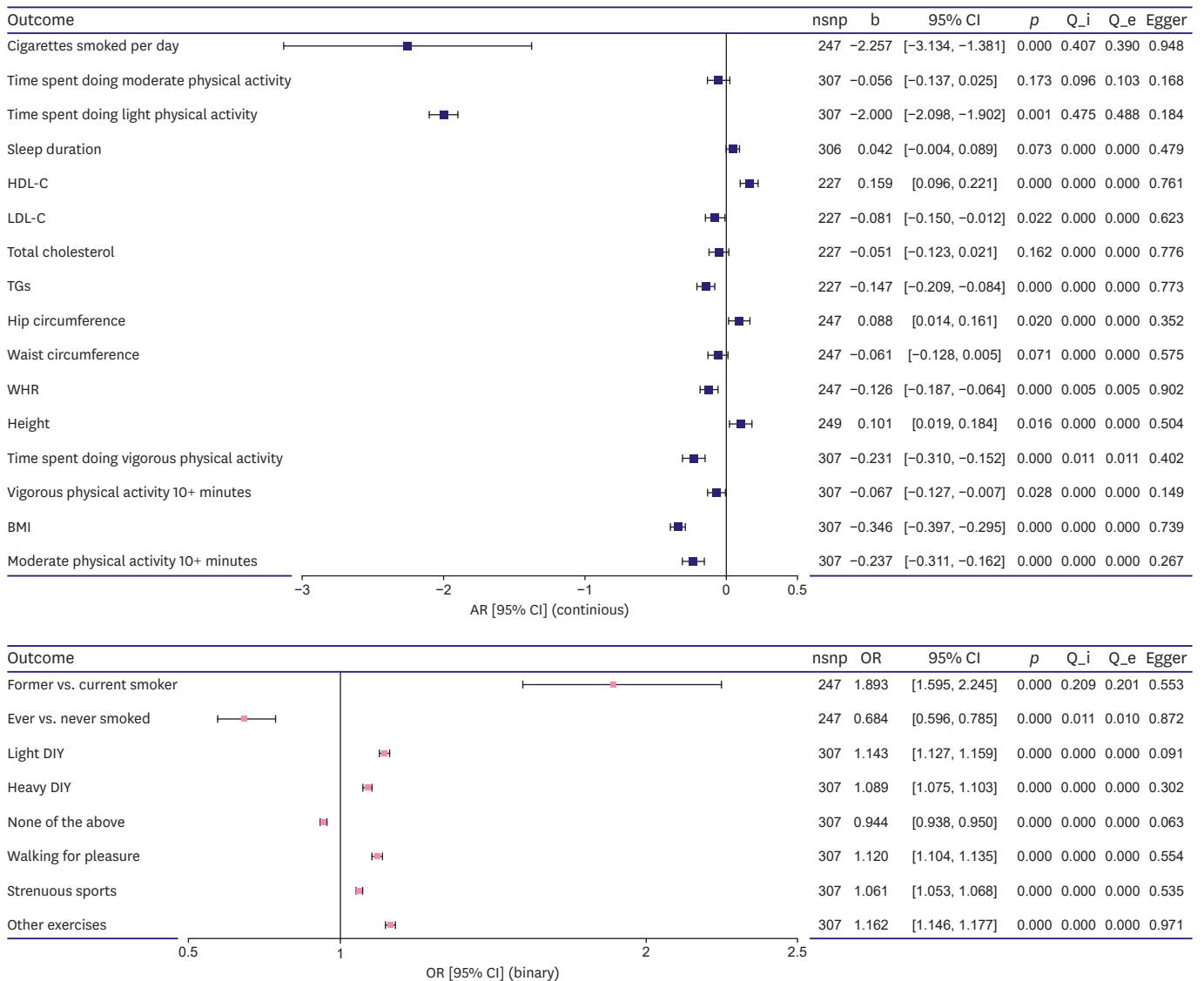| Outcome		nsnp	OR	95% CI	p	Q_i	Q_e	Egger
Former vs. current smoker		247	1.893	[1.595, 2.245]	0.000	0.209	0.201	0.553
Ever vs. never smoked		247	0.684	[0.596, 0.785]	0.000	0.011	0.010	0.872
Light DIY		307	1.143	[1.127, 1.159]	0.000	0.000	0.000	0.091
Heavy DIY		307	1.089	[1.075, 1.103]	0.000	0.000	0.000	0.302
None of the above		307	0.944	[0.938, 0.950]	0.000	0.000	0.000	0.063
Walking for pleasure		307	1.120	[1.104, 1.135]	0.000	0.000	0.000	0.554
Strenuous sports		307	1.061	[1.053, 1.068]	0.000	0.000	0.000	0.535
Other exercises		307	1.162	[1.146, 1.177]	0.000	0.000	0.000	0.971

OR [95% CI] (binary)

Figure 3. Causal effects of the level of education on 25 risk factors of breast cancer.
CI = confidence interval; Q_i = p-value of Q statistics in inverse variance weighted method; Q_e = p-value of Q statistics in MR-Egger regression method; egger = p-value of the intercept in the MR-Egger regression; MR = Mendelian randomization; HDL-C = high-density lipoprotein cholesterol; LDL-C = low-density lipoprotein cholesterol; TG = triglyceride; WHR = waist-to-hip ratio; BMI = body mass index; AR = attributable risk; DIY = do-it-yourself; OR = odds ratio.

ER-negative breast cancer (light DIY activities: OR, 0.923; MP, 6.885%; walking for pleasure: OR, 0.867; MP, 12.173%; engaging in strenuous sports: OR, 0.937; MP, 5.554%). On the contrary, smoking and heavy DIY had indirect effects on the risk of ER-negative breast cancer; that is, education increases the risk of ER-negative breast cancer through smoking and performance of heavy DIY activities (ever vs. never smoked: OR, 1.130; MP, 10.442%; former vs. current smoker: OR, 1.091; MP, 7.420%; heavy DIY activities: OR, 1.036; MP, 3.033%).

Therefore, we also calculated the indirect effects of the 20 risk factors on the association between education and ER-positive breast cancer. For continuous mediators, increased education levels reduced the risk of ER-positive breast cancer through the effects of engaging in vigorous physical activities (OR, 0.977; MP, 1.94%), HDL-cholesterol level (OR, 1.030; MP, 2.47%), TG level (OR, 1.032; MP, 2.58%), WHR (OR, 1.031; MP, 2.50%), height

Table 3. Indirect effects through each mediator from educational attainment to the development of 2 breast cancer subtypes and their mediation proportions

Mediators	ER⁻			ER⁺		
	Indirect effect		Mediation proportion	Indirect effect		Mediation proportion
	log(OR)	OR		log(OR)	OR	
Continue mediators						
HDL-C	0.023	1.023	1.962%	0.030	1.030	2.47%
LDL-C	−0.004	0.996	0.331%	−0.006	0.994	0.51%
TGs	0.022	1.022	1.894%	0.031	1.032	2.58%
Hip circumference	0.000	1.000	0.001%	0.021	1.021	1.74%
WHR	0.020	1.021	1.743%	0.030	1.031	2.50%
Cigarettes smoked per day	0.002	1.002	0.149%	−0.007	0.993	0.55%
Height	0.001	1.001	0.043%	0.027	1.028	2.25%
Time spent doing vigorous physical activity	0.030	1.030	2.544%	−0.023	0.977	1.94%
Vigorous physical activity 10+ minutes	0.010	1.010	0.882%	0.013	1.013	1.11%
BMI	0.071	1.073	6.050%	0.069	1.072	5.75%
Moderate physical activity 10+ minutes	0.031	1.032	2.664%	0.006	1.006	0.48%
Time spent doing light physical activity	0.000	1.000	0.015%	0.000	1.000	0.03%
Binary mediators						
Ever vs never smoked	0.122	1.130	10.442%	−0.066	0.936	5.447%
Former vs current smoker	0.087	1.091	7.420%	0.073	1.076	6.087%
Light DIY	−0.081	0.923	6.885%	0.019	1.019	1.595%
Heavy DIY	0.035	1.036	3.033%	−0.074	0.929	6.093%
None of the above	−0.112	0.894	9.576%	−0.102	0.903	8.412%
Walking for pleasure	−0.142	0.867	12.173%	−0.159	0.853	13.147%
Strenuous sports	−0.065	0.937	5.554%	−0.076	0.927	6.263%
Other exercises	−0.178	0.837	15.189%	−0.123	0.885	10.157%

ER⁺ = estrogen receptor-positive breast cancer risk; ER⁻ = estrogen receptor-negative breast cancer risk; OR = odds ratio; HDL-C = high-density lipoprotein cholesterol; LDL-C = low-density lipoprotein cholesterol; TG = triglyceride; WHR = waist-to-hip ratio; BMI = body mass index; DIY = do-it-yourself.

measurement (OR, 1.028; MP, 2.25%), and BMI (OR, 1.072; MP, 5.75%) as these factors may pose potential hazards to the protective pathway from the low of educational levels to the development of ER-positive breast cancer; that is, the indirect effects of these mediators increased the risk of ER-positive breast cancer. For binary mediators, increased education levels increased the risk of ER-positive breast cancer through the effect of performing light DIY activities. On the contrary, physical activity is a protective mediator of the pathway from educational attainment to the development of ER-positive breast cancer and accounts for a large MP (heavy DIY: OR, 0.929; MP, 6.093%; walking for pleasure: OR, 0.853; MP, 13.147%; performance of strenuous sports: OR, 0.927; MP, 6.263%; other exercise: OR, 0.885; MP, 10.157%). Other mediators only explain a small part of the causal effect of education on the risk of developing ER-positive breast cancer, and their MP is less than 1%.

DISCUSSION

Our study showed that every increase in educational level decreased the risk of ER-negative breast cancer by 23% (OR, 0.77; 95% CI, 0.6–0.984; p = 0.004). However, no causal association was found between overall breast cancer and the risk of ER-positive breast cancer, which was consistent with the results of the sensitivity analyses. Lipid profile, obesity, smoking, and physical activities were identified as mediators in the causal pathway from educational attainment to the development of breast cancer. Education level may affect the risk of developing ER-positive breast cancer through several mediators, but the sum of direct and indirect effects through each mediator is close to null. Physical activities can help improve the protective effect of education against breast cancer, with relatively large MPs. Education increases the risk of ER-negative breast cancer through the effects of HDL levels,

TG levels, height measurement, WHR, BMI, and smoking, with relative medium MPs. Other mediators including LDL, hip circumference, number of cigarettes smoked per day, time spent performing light physical activities, and engaging in vigorous physical activities for > 10 minutes explain a small part of the causal effect of education on the risk of developing breast cancer, and their MP are approximately 1%.

A large meta-analysis including more than 10 million women found that a high degree of education may be associated with a higher risk of breast cancer. In addition, menopausal age, alcohol consumption, and hormone therapy may mediate this causal effect to a certain extent [5]. However, this may be hampered by the underlying sources of bias (e.g., unmeasured confounding and reverse causation). A cohort study of 3,092 individuals born in Limache Hospital between 1974 and 1978 showed that poor education may be associated with a poor lipid profile in women [21]. Consistent with this study, a two-sample MR study conducted in 400,000 participants [22] showed that increased LDL-cholesterol (LDL-C) levels were associated with a higher risk of breast cancer (OR, 1.09; 95% CI, 1.02–1.18; $p = 0.020$) and ER-positive breast cancer (OR, 1.14; 95% CI, 1.05–1.24; $p = 0.004$). Individuals with genetically higher HDL-cholesterol levels were at an increased risk of developing ER-positive breast cancer (OR, 1.13; 95% CI, 1.01–1.26; $p = 0.037$). Higher HDL-cholesterol and lower TG levels were found to be determinants of ER-negative breast cancer. However, HDL-cholesterol, LDL-C, and TG levels were not significantly associated with either overall breast cancer risk or ER-negative breast cancer risk. Rodrigues Dos Santos et al. [23] indicated that ER-negative tumors are particularly sensitive to elevated cholesterol levels and, given the increasing appreciation of the role of liver X receptor-alpha (LXR) signaling in breast cancer, potentially explain why ER-negative disease is more likely to be altered by cholesterol-lowering interventions than ER-positive disease. A limitation of their study is the lack of stratification of women by menopausal status. Endocrine changes during menopause may alter the lipid composition and the interaction with breast tissue. For example, a meta-analysis of observational studies found a negative association between HDL cholesterol and breast cancer only in postmenopausal women, but not in premenopausal women [24]. A further limitation of this study is that the effects of age at menopause and hormone therapy were not considered. Hence, further studies are warranted to investigate the role of menopausal status in the causal pathway from education attainment to the development of breast cancer.

A cross-sectional study indicated that low levels of education are independently associated with obesity [25]. Consistent with our results, another MR study suggested that increased BMI could lower the breast cancer survival in ER-positive breast cancer patients [26]. However, their results indicated that BMI had no causal effect on ER-negative breast cancer. A limitation of their analysis is that there might be a selection bias from the genetic variants associated with these confounders in the subpopulation of breast cancer patients. This is due to the conditioning on a collider and is referred to as selection bias; this finding indicates the need to select a representative population for MR analysis. A simulation study found that selection bias significantly affects the estimation of the causal effect and the type 1 error rate only when the selection effect is large [27]. LDL-C, obesity, WHR, and waist circumference are associated with the incidence and survival of breast cancer [23,28] and the clinically recommended diets/lifestyle changes that lower LDL-C and protect against breast cancer and relapse, particularly in the hormone receptor-negative setting [29,30]. Pharmacological manipulation of LDL-C with lipophilic statins improves breast cancer survivorship, specifically reducing early (< 4 years) relapse events, a feature typical of an ER-negative disease [23].

In accordance with our results, another MR study indicated that accelerometer-measured physical activity was negatively associated with the risk of breast cancer. Multiple biological mechanisms have been proposed to explain the potential beneficial effects of physical activity on breast cancer development. Physical activity can reduce the levels of circulating insulin and insulin-like growth factor, promote cell proliferation in breast tissues, and prevent cancer development in these areas. High levels of physical activity also reduced the circulating levels of estradiol and increased the levels of sex hormone-binding globulin [31], which are risk factors for breast cancer. The significant associations shown for ER cancers instead of ER$^+$-positive cancers suggest that non-hormonal mechanisms may also play a role in the protective effect of physical activity [32]. However, only a few studies have investigated the causal relationship between the level of education and physical activity. Hildreth et al. [32] suggested that ER-positive and ER-negative breast cancers may share some risk factors, but not others, because of the inconsistent results among different studies. Previous research has shown that hormone-related factors, such as age at menarche, parity, and age at menopause, tend to be associated with receptor-positive (ER$^+$/PR$^+$) breast cancer, whereas family history of breast cancer and cigarette smoking have been associated with receptor-negative (ER$^-$/PR$^-$) breast cancer (7–16). These findings suggest that breast cancer does not represent a single phenotype (i.e., that it is not a homogeneous disease) but rather a heterogeneous set of diseases with perhaps different genetic and environmental determinants.

Our study had several important advantages. We conducted an MR study to investigate the causal relationship between the level of education and breast cancer. Participants were grouped according to their randomly assigned genotypes, similar to randomized control trials. The MR method avoids the interference of reverse causation and potential confounding factors that are common in conventional observational studies. The large sample size of summary datasets improves the statistical power and estimates the causal effect with high precision. The strong instrumental variables (F statistics > 10) [33] compensated for the weak instrumental bias. We also unlocked the mechanism in the causal pathway from educational attainment to the development of breast cancer. We calculated the MPs of pathways from each mediator and divided the mediators into 3 groups: large, medium, and small. We also revealed that the inconsistent direction of indirect effects was the same as the total effect of education on the risk of breast cancer. Another advantage of our study is that we focused on the causal effects of education on the risk of breast cancer and the mediators instead of associations.

Our study has several limitations. First, all the participants included in our study were of European descent. Thus, it is unknown whether our findings can be applied to other ethnicities. In addition, the InSIDE assumption in the MR-Egger test remains a limitation. In the InSIDE assumption, the effect of genetic variants on exposure is independent of the direct effects of genetic variants on the outcome, which is difficult to evaluate. Therefore, further studies are warranted to investigate the role of menopausal status in the causal pathway between the level of education and breast cancer.

In conclusion, our present Mendelian randomization study provided strong evidence to suggest that higher educational attainment played a causal role in lowering the risk of breast cancer. A low level of education is a causal risk factor in the development of breast cancer, as it is associated with a poor lipid profile, anthropometric measurements, smoking, and types of physical activity.

SUPPLEMENTARY MATERIALS

Supplementary Table 1
Summary of the results of systematic literature review

Click here to view

Supplementary Data 1
Mendelian randomization and calculation of mediation proportion

Click here to view

Supplementary Figure 1
Forest plot of the meta-analysis (OR).

Click here to view

Supplementary Figure 2
Forest plot of the meta-analysis (RR).

Click here to view

Supplementary Figure 3
Forest plot of the meta-analysis (HR).

Click here to view

Supplementary Figure 4
Forest plot of SNPs associated with education and their risk of ER-negative breast cancer (combined).

Click here to view

Supplementary Figure 5
Scatter plot of SNPs associated with education and their risk of ER-negative breast cancer (combined).

Click here to view

Supplementary Figure 6
Leave-one-out plot of SNPs associated with education and their risk of ER-negative breast cancer (combined).

Click here to view

Supplementary Figure 7

Forest plot of SNPs associated with education and their risk of ER-negative breast cancer (OncoArray).

Click here to view

Supplementary Figure 8

Scatter plot of SNPs associated with education and their risk of ER-negative breast cancer (OncoArray).

Click here to view

Supplementary Figure 9

Leave-one-out plot of SNPs associated with education and their risk of ER-negative breast cancer (OncoArray).

Click here to view

Supplementary Figure 10

Forest plot of SNPs associated with education and their risk of ER-negative breast cancer (iCOGS).

Click here to view

Supplementary Figure 11

Scatter plot of SNPs associated with education and their risk of ER-negative breast cancer (iCOGS).

Click here to view

Supplementary Figure 12

Leave-one-out plot of SNPs associated with education and their risk of overall breast cancer (iCOGS).

Click here to view

Supplementary Figure 13

Forest plot of SNPs associated with education and their risk of ER-negative breast cancer (GWAS).

Click here to view

Supplementary Figure 14

Scatter plot of SNPs associated with education and their risk of ER-negative breast cancer (GWAS).

Click here to view

Supplementary Figure 15
Leave-one-out plot of SNPs associated with education and their risk of ER-negative breast cancer (GWAS).

Click here to view

Supplementary Figure 16
Causal estimates of education on potential mediators (binary).

Click here to view

Supplementary Figure 17
Causal estimates of education on potential mediators (continue).

Click here to view

Supplementary Figure 18
Causal estimates of potential mediators on ER-negative breast cancer (combined).

Click here to view

Supplementary Figure 19
Causal estimates of potential mediators on ER-negative breast cancer (GWAS).

Click here to view

Supplementary Figure 20
Causal estimates of potential mediators on ER-negative breast cancer (iCOGS).

Click here to view

Supplementary Figure 21
Causal estimates of potential mediators on ER-negative breast cancer (OncoArray).

Click here to view

REFERENCES

1. Harbeck N, Gnant M. Breast cancer. Lancet 2017;389:1134-50.
   PUBMED | CROSSREF

2. Global Burden of Disease Cancer CollaborationFitzmaurice C, Allen C, Barber RM, Barregard L, Bhutta ZA, et al. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 32 cancer groups, 1990 to 2015: a systematic analysis for the global burden of disease study. JAMA Oncol 2017.3:524-48.
   PUBMED | CROSSREF

3. DeSantis C, Ma J, Bryan L, Jemal A. Breast cancer statistics, 2013. CA Cancer J Clin 2014;64:52-62.
   PUBMED | CROSSREF

4. DeSantis C, Howlader N, Cronin KA, Jemal A. Breast cancer incidence rates in U.S. women are no longer declining. Cancer Epidemiol Biomarkers Prev 2011;20:733-9.
   PUBMED | CROSSREF

5. Dong JY, Qin LQ. Education level and breast cancer incidence: a meta-analysis of cohort studies. Menopause 2020;27:113-8.
    PUBMED | **CROSSREF**

6. Menvielle G, Kunst AE, van Gils CH, Peeters PH, Boshuizen H, Overvad K, et al. The contribution of risk factors to the higher incidence of invasive and *in situ* breast cancers in women with higher levels of education in the European prospective investigation into cancer and nutrition. Am J Epidemiol 2011;173:26-37.
    PUBMED | **CROSSREF**

7. Goldberg M, Calderon-Margalit R, Paltiel O, Abu Ahmad W, Friedlander Y, Harlap S, et al. Socioeconomic disparities in breast cancer incidence and survival among parous women: findings from a population-based cohort, 1964–2008. BMC Cancer 2015;15:921.
    PUBMED | **CROSSREF**

8. Carlsen K, Høybye MT, Dalton SO, Tjønneland A. Social inequality and incidence of and survival from breast cancer in a population-based study in Denmark, 1994–2003. Eur J Cancer 2008;44:1996-2002.
    PUBMED | **CROSSREF**

9. Hajian-Tilaki K, Kaveh-Ahangar T, Hajian-Tilaki E. Is educational level associated with breast cancer risk in Iranian women? Breast Cancer 2012;19:64-70.
    PUBMED | **CROSSREF**

10. Bjerkaas E, Parajuli R, Engeland A, Maskarinec G, Weiderpass E, Gram IT. Social inequalities and smoking-associated breast cancer - results from a prospective cohort study. Prev Med 2015;73:125-9.
    PUBMED | **CROSSREF**

11. Hussain SK, Altieri A, Sundquist J, Hemminki K. Influence of education level on breast cancer risk and survival in Sweden between 1990 and 2004. Int J Cancer 2008;122:165-9.
    PUBMED | **CROSSREF**

12. Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for Mendelian randomization. Stat Methods Med Res 2017;26:2333-55.
    PUBMED | **CROSSREF**

13. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol 2013;37:658-65.
    PUBMED | **CROSSREF**

14. Zhou H, Zhang Y, Liu J, Yang Y, Fang W, Hong S, et al. Education and lung cancer: a Mendelian randomization study. Int J Epidemiol 2019;48:743-50.
    PUBMED | **CROSSREF**

15. Gill D, Efstathiadou A, Cawood K, Tzoulaki I, Dehghan A. Education protects against coronary heart disease and stroke independently of cognitive function: evidence from Mendelian randomization. Int J Epidemiol 2019;48:1468-77.
    PUBMED | **CROSSREF**

16. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat Genet 2018;50:1112-21.
    PUBMED | **CROSSREF**

17. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat Genet 2013;45:353-61.
    PUBMED | **CROSSREF**

18. Shungin D, Winkler TW, Croteau-Chonka DC, Ferreira T, Locke AE, Mägi R, et al. New genetic loci link adipose and insulin biology to body fat distribution. Nature 2015;518:187-96.
    PUBMED | **CROSSREF**

19. Hou L, Li H, Si S, Yu Y, Sun X, Liu X, et al. Exploring the causal pathway from bilirubin to CVD and diabetes in the UK biobank cohort study: observational findings and Mendelian randomization studies. Atherosclerosis 2021;320:112-21.
    PUBMED | **CROSSREF**

20. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. Genet Epidemiol 2016;40:597-608.
    PUBMED | **CROSSREF**

21. Lara M, Amigo H. Association between education and blood lipid levels as income increases over a decade: a cohort study. BMC Public Health 2018;18:286.
    PUBMED | **CROSSREF**

22. Nowak C, Ärnlöv J. A Mendelian randomization study of the effects of blood lipids on breast cancer risk. Nat Commun 2018;9:3957.
    PUBMED | **CROSSREF**

23. Rodrigues Dos Santos C, Fonseca I, Dias S, Mendes de Almeida JC. Plasma level of LDL-cholesterol at diagnosis is a predictor factor of breast tumor progression. BMC Cancer 2014;14:132.
PUBMED | CROSSREF

24. Ni H, Liu H, Gao R. Serum lipids and breast cancer risk: a meta-analysis of prospective cohort studies. PLoS One 2015;10:e0142669.
PUBMED | CROSSREF

25. Sánchez CN, Maddalena N, Penalba M, Quarleri M, Torres V, Wachs A. Relationship between level of education and overweight in outpatients. A transversal study. Medicina (B Aires) 2017;77:291-6.
PUBMED

26. Guo Q, Burgess S, Turman C, Bolla MK, Wang Q, Lush M, et al. Body mass index and breast cancer survival: a Mendelian randomization analysis. Int J Epidemiol 2017;46:1814-22.
PUBMED | CROSSREF

27. Gkatzionis A, Burgess S. Contextualizing selection bias in Mendelian randomization: how bad is it likely to be? Int J Epidemiol 2019;48:691-701.
PUBMED | CROSSREF

28. dos Santos CR, Domingues G, Matias I, Matos J, Fonseca I, de Almeida JM, et al. LDL-cholesterol signaling induces breast cancer proliferation and invasion. Lipids Health Dis 2014;13:16.
PUBMED | CROSSREF

29. Link LB, Canchola AJ, Bernstein L, Clarke CA, Stram DO, Ursin G, et al. Dietary patterns and breast cancer risk in the California Teachers Study cohort. Am J Clin Nutr 2013;98:1524-32.
PUBMED | CROSSREF

30. World Cancer Research Fund (WCRF)/American Institute for Cancer Research (AICR). *Continuous Update Project Report: Diet, Nutrition, Physical Activity and Cancer: A Global Perspective*. London: World Cancer Research Fund International/American Institute for Cancer Research; 2017.

31. Liedtke S, Schmidt ME, Becker S, Kaaks R, Zaineddin AK, Buck K, et al. Physical activity and endogenous sex hormones in postmenopausal women: to what extent are observed associations confounded or modified by BMI? Cancer Causes Control 2011;22:81-9.
PUBMED | CROSSREF

32. Hildreth NG, Kelsey JL, Eisenfeld AJ, LiVolsi VA, Holford TR, Fischer DB. Differences in breast cancer risk factors according to the estrogen receptor level of the tumor. J Natl Cancer Inst 1983;70:1027-31.
PUBMED

33. Hou L, Xu M, Yu Y, Sun X, Liu X, Liu L, et al. Exploring the causal pathway from ischemic stroke to atrial fibrillation: a network Mendelian randomization study. Mol Med 2020;26:7.
PUBMED | CROSSREF