Research paper

# Physical activities and breast cancer□A Mendelian Randomization Study

**Keywords**
Mendelian randomization, Physical activities, breast cancer disease, causal estimates

**Abstract**

Introduction
Previous research suggests a potential association between physical activity (PA) and breast cancer (BC), but the causal relationship remains uncertain. The aim of this study was to explore the causal relationship between PA and BC through Mendelian randomization (MR) analysis.

Material and methods
Genome-wide association studies utilizing data from the UK Biobank baseline were employed to analyze PA phenotypes, encompassing 460,376 participants. Summary data for BC, comprising 122,977 cases and 105,974 controls, were obtained from the Breast Cancer Association Consortium (BCAC). The cases were further categorized based on estrogen receptor status into estrogen receptor-positive (ER+) BC and estrogen receptor-negative (ER−) BC. The inverse variance weighted method was employed as the primary approach for two-sample MR. Additionally, MR-PRESSO (MR-Pleiotropy RESidual Sum and Outlier) method was utilized to eliminate outliers. Tests for heterogeneity and pleiotropy were conducted to enhance result accuracy. Furthermore, multivariable Mendelian randomization was performed, adjusting for potential confounders to ensure result stability.

Results
Mendelian randomization (MR) analysis was employed to assess the causal link between PA and BC. Two-sample MR analysis revealed a genetic prediction indicating that walking for pleasure was associated with decreased risk of (ER+) BC (odds ratio [OR]=0.302, 95% CI =-2.257– -0.137, p=0.027), other physical activities were not significantly correlated with BC, (ER+) BC and (ER−) BC. These findings remained reliable and consistent in the sensitivity analysis, including Cochran’s Q and MR-Egger regression.

Conclusions
Our findings suggest that engaging in leisure walking is associated with a reduced risk of (ER+) BC.

# Physical activities and breast cancer: A Mendelian Randomization Study


## Abstract

**Background:** Previous research suggests a potential association between physical activity (PA) and breast cancer (BC), but the causal relationship remains uncertain. The aim of this study was to explore the causal relationship between PA and BC through Mendelian randomization (MR) analysis.

**Methods:** Genome-wide association studies utilizing data from the UK Biobank baseline were employed to analyze PA phenotypes, encompassing 460,376 participants. Summary data for BC, comprising 122,977 cases and 105,974 controls, were obtained from the BC Association Consortium. The cases were further categorized based on estrogen receptor status into estrogen receptor-positive breast cancer (ER+ BC) and estrogen receptor-negative breast cancer (ER− BC). The inverse variance weighted method was employed as the primary approach for two-sample MR. Additionally, MR-PRESSO (MR-Pleiotropy RESidual Sum and Outlier) method was utilized to eliminate outliers. Tests for heterogeneity and pleiotropy were conducted to enhance result accuracy. Furthermore, multivariable Mendelian randomization was performed, adjusting for potential confounders to ensure result stability.

**Results:** Mendelian randomization (MR) analysis was employed to assess the causal link between PA and BC. Two-sample MR analysis revealed a genetic prediction indicating that walking for pleasure was associated with decreased risk of ER+ BC (odds ratio [*OR*]=0.302, 95% *CI* =0.105-0.872, *p*=0.027), other physical activities were not significantly correlated with BC, ER+ BC and ER− BC. These findings remained reliable and consistent in the sensitivity analysis, including Cochran's Q and MR-Egger regression. Furthermore, reverse MR analysis suggested that BC did not exert a notable impact on PA.

**Conclusion:** Our findings suggest that engaging in leisure walking is associated with a reduced risk of ER+ BC. Nevertheless, additional research is warranted to comprehensively elucidate the underlying mechanisms and strengthen the causal relationship.

**Keywords:** Physical activities; breast cancer disease; causal estimates; Mendelian randomization

## 1.Introduction

Breast cancer (BC) ranks as the most prevalent cancer among women, accounting for 30% of all cases and contributing to 15% of deaths in 2022[1, 2]. It has been highlighted in recent reports that BC is recognized as the leading malignancy among the cancer-related deaths of women around the world. Furthermore, the incidence number and rate of mortality has been constantly rising since the twenty-first century[3]. In order to lower the increasing burden of BC[4], a number of researchers have been actively investigating the potential influence and risk factors for BC so as to reduce the incidence rate of BC[5, 6].

Physical activity (PA) and exercise are considered to exert positive effect on the handling of various chronic diseases, especially with regard to the prevention and treatment[7, 8]. Individuals reporting higher levels of PA tended to exhibit better overall health[9, 10]. However, the engagement of PA is much different from person to person. And culture as well as economy has an impact on that as part of the environmental factors. Multiple research articles have provided the evidence that the predisposition of humans to exercise is associated with their genetic factors[11, 12]. Several epidemiological studies have investigated the relationship between physical activities and breast cancer with conflicting findings[13-15]. While some prospective cohort studies have suggested a link between physical activities and reduced breast cancer risk[16, 17], other studies have shown that physical activities are not correlated to the risk of breast cancer [18, 19]. Taking the limitations of observational studies into account, random and systematic errors affect the validity of the findings above as a result, including potential selection bias, effects of cohort design bias, limited sample size, missed follow-ups as well as the presence of reverse causality between outcomes and exposure. In addition, ethical issues, cost, as well as long follow-ups restrict randomized controlled trials. It remains unknown whether physical activities play a causal role in breast cancer. Besides, it is difficult to determine the specific distinctions about the duration, intensity, and type of exercise. In particular, the best solution for the types of exercise that patients with breast cancer can choose is even harder to decide. Thus, a two-sample Mendelian randomization (MR) method was utilized in this study to discover potential causal link between different types of physical activities and breast cancer.

In the MR analysis, single nucleotide polymorphisms (SNPs) which have strong relationship with exposure, for example heavy physical activity, are considered as instrumental variables (IVs) to estimate the causal effect with outcome (e.g., BC). MR is a 'natural' RCT that makes use of the random distribution of genetic variants with influence on exposure[20]. Those SNPs that are strongly linked with confounders will be eliminated before performing the MR analysis in order to remove the effect of confounding factors. Reverse MR analysis can exclude potential reverse causal effect between exposures and outcomes. In this study, five types of physical activity with various intensities were analyzed to investigate their association with BC. Furthermore, bilateral MR was performed by using datasets from genome-wide association study (GWAS) to examine the causal link between physical activity and BC.

## 2.Materials and methods

### 2.1 Study design

A two-sample Mendelian randomization (MR) study was designed to estimate the potential causal link between PA and BC. The single nucleotide polymorphisms (SNPs) were selected as IVs and stick to three essential premises as follows[21]: (1) SNPs should be intensely linked to PA as exposure; (2) SNPs should not be linked to confounding factors; and (3) SNPs should not be linked to BC as outcome directly (**Figure 1**).

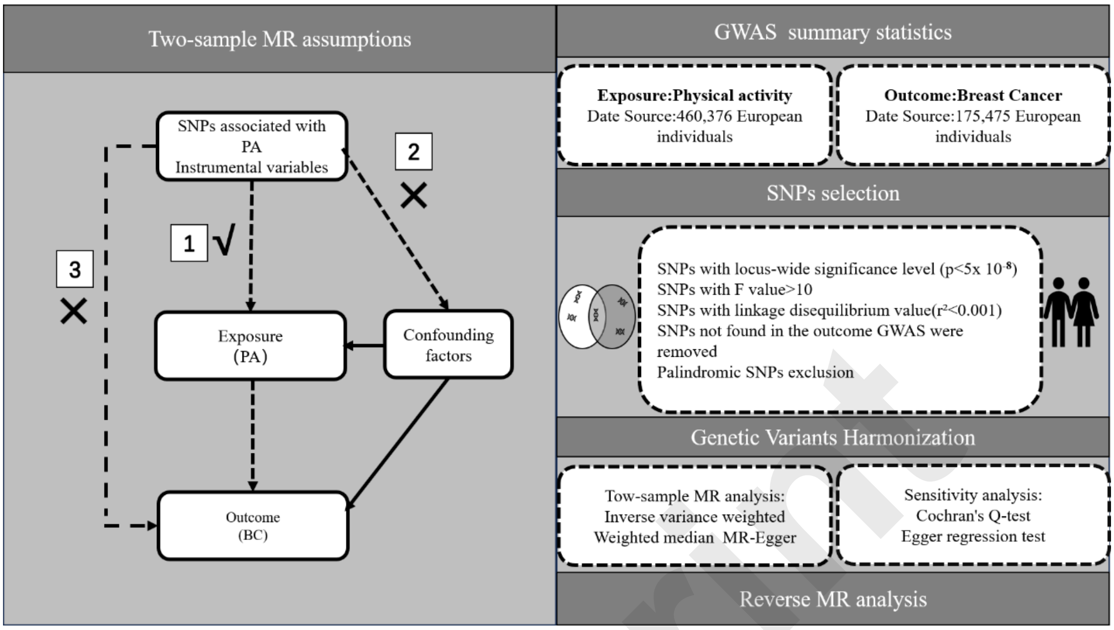



**Figure 1**. Study design to assess the correlation between physical activities and risk of breast cancer based on the assumptions of bidirectional Mendelian randomization.



### 2.2 GWAS summary statistics

The summary statistics for physical activity of various types were acquired and extracted from an online public database (IEU Open GWAS Project [https://gwas.mrcieu.ac.uk/](https://gwas.mrcieu.ac.uk/)). In this study, five types of physical activity corresponding to different intensities were selected from the database and utilized to investigate the causal association with breast cancer: heavy DIY, light DIY, strenuous sports, walking for pleasure, and other exercises. The data were collected by asking participants to fill in questionnaires using touchscreen; the participants were provided with the different options above and asked to choose the one that they had been involved most in the last month. Furthermore, the survey included 497,174 European participants of both males and females. The physical activity assessment in this study was validated by 4 instances, including 497,235 participants, thus ensuring a large enough sample size for reliable results; specifically, instance 0 was the initial assessment visit (2006 – 2010), instance 1 was the first repeat assessment visit (2012 – 13), instance 2 was the imaging visit (2014), and the last instance was the first imaging visit (2019).

The summary data for BC which includes 122,977 cases and 105,974 controls was extracted from the Breast Cancer Association Consortium. Based on the estrogen receptor status, the cases were further classified into two categories: estrogen receptor-positive Breast cancer (ER+ BC) and estrogen receptor-negative Breast cancer (ER− BC). **Table 1** presents details of the exposure and outcomes.

**Table 1** Detailed information on the exposure and outcomes.

| **Exposure/Outcome** | ncase | ncontrol | Sample size | Ancestry | MRC-IEU ID |
|---|---|---|---|---|---|
| Heavy DIY (eg: weeding, lawn mowing, carpentry, digging) | 197,006 | 263,370 | 460,376 | European | ukb-b-13184 |
| Light DIY (eg: pruning, watering the lawn) | 236,244 | 224,132 | 460,376 | European | ukb-b-11495 |
| Strenuous sports | 47,468 | 412,908 | 460,376 | European | ukb-b-7663 |
| Walking for pleasure (not as a means of transport) | 329,755 | 130,621 | 460,376 | European | ukb-b-7337 |
| Other exercises (eg: swimming, cycling, keep fit, bowling) | 222,470 | 237,906 | 460,376 | European | ukb-b-8764 |
| Breast cancer | 122,977 | 105,974 | 228,951 | European | ieu-a-1126 |
| ER+ BC | 69,501 | 105,974 | 175,475 | European | ieu-a-1127 |
| ER- BC | 21,468 | 105,974 | 127,442 | European | ieu-a-1128 |

ER+ BC: estrogen receptor status into estrogen receptor-positive; ER− BC: estrogen receptor-negative.

### 2.3 Ethical approval

All summary-level datasets in our study were obtained from de-identified public data/studies. Ethical approval and informed consent were previously obtained from the ethics committee. Thus, the requirement for ethical approval was waived for this study.

**2.4 SNPs selection**

Firstly, we conducted a screening process to identify SNPs that were highly correlated with exposure at a genome-wide significance level ($p < 5\times10^{-8}$). Secondly, we implemented a criterion ($r^2 < 0.001$, kb=10000) to choose SNPs that were free from dependence on linkage disequilibrium (LD). Thirdly, we excluded SNPs that were not present in the BC dataset and palindromic SNPs which have the potential to introduce bias. All of the SNPs for instrumental variables were uploaded to PhenoScanner to identify confounding SNPs associated with BC. Based on the assumption of the MR analysis, SNPs used as instrumental variables should be strongly associated with exposure. Subsequently, we ensured the harmonization of exposure and outcome data, confirming that the effect of the SNP on the exposure corresponded to the same allele as its effect on the outcome. Following this, we assessed the possibility of weak instrumental bias by calculating F-statistics, and excluded SNPs with F-statistics less than 10. The F statistic was calculated using the formula $F = beta^2/se^2$. Finally, we employed the MR-PRESSO method to identify outlier SNPs. After removing the outliers, the remaining SNPs were utilized for subsequent MR analysis. A flowchart illustrating the selection process is provided in **Figure 1**.



**2.5 Two-sample Mendelian analysis**

Three popular MR methods were employed to assess causal effects: inverse variance weighted (IVW), weighted median and MR-Egger[22, 23]. IVW, a reliable and robust MR method in the absence of horizontal pleiotropy[24], combines the Wald estimates of individual SNP to derive overall estimates of the effect of physical activities on breast cancer risk. Consequently, the IVW method is broadly acknowledged as the most effective approach to assess causality. Odds ratios (ORs) were utilized to express the effects of physical activities on BC risk. If the result of the IVW method is significant ($p < 0.05$), it can be considered positive even if other methods yield nonsignificant results, provided that the ORs of those methods line up in the identical direction without heterogeneity or pleiotropy. Two types of IVW approaches, namely the fixed and random effect model, were employed to account for existing heterogeneity. Cochran's Q test was used to assess the heterogeneity in the IVW method and MR-Egger regression, with a P-value < 0.05 considered statistically significant[25]. Unlike IVW, the MR-Egger method includes an intercept term designed to test for horizontal pleiotropy. A non-zero intercept term indicates that not all genetic variants are valid instruments, thereby biasing IVW estimates. When the instrument strength independent of direct effect (InSIDE) assumption is met, the MR–Egger method can offer an approximation of the causal impact of horizontal pleiotropy[26]. The weighted median method offers a robust effect estimate, even in the presence of unbalanced horizontal pleiotropy (e.g., when 50% of instrumental SNPs are invalid). Finally, the MR-PRESSO method encompasses three detection functions[27, 28]: horizontal pleiotropic detection, horizontal pleiotropic correction (after outlier removal), as well as assessment of differences in the results of causality estimation before and after correction.

**2.6 Statistical analysis`**

Heterogeneity was assessed by employing Cochran's Q test[29], where a p-value > 0.05 indicated the absence of heterogeneity. The MR-Egger regression test was utilized to identify horizontal pleiotropy, where a zero-intercept suggests the absence of pleiotropy (p > 0.05).

**2.7 Reverse MR analysis**

To explore the potential causal relationship between BC and PA, a reverse MR analysis was carried out, wherein BC served as the exposure and PA as the outcome, employing SNPs associated with BC as IVs.

All statistical analyses were conducted using R software (version 4.2.3) with the "TwoSampleMR" (version 0.5.6), "MRPRESSO" (version 1.0), and "MendelianRandomization" (version 0.7.0) packages.

## 3.Results

### 3.1 The MR analysis results.

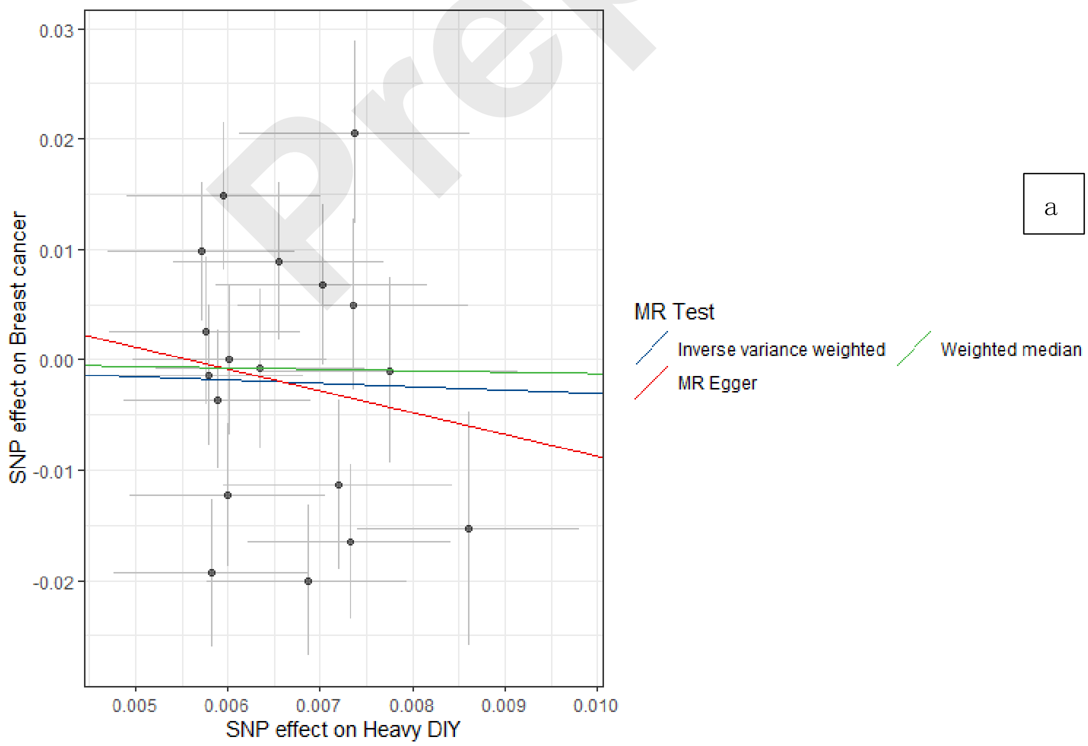
The results of the MR analysis for the three methods are presented in Table 2 for physical activities, BC, ER+ BC and ER- BC. The MR estimates suggested that walking for pleasure conferred a protective effect against ER+ BC (odds ratio [*OR*]=0.302, 95% *CI* =-2.257– -0.137, *p*=0.027). No causal effect was observed for the other four types of physical activities on BC, ER+ BC and ER- BC. Scatter plots depicting the MR analysis of the causal effect of physical activity on BC, ER+ BC, and ER- BC are presented in Figures 2, 3, and 4, respectively. Cochran's Q and MR-Egger regression analyses revealed no evidence of heterogeneity or horizontal pleiotropy affecting the stability of the results. Hence, drawing from the IVW findings ($p < 0.05$), we can infer the presence of a causal relationship between leisurely walking and ER+ BC.

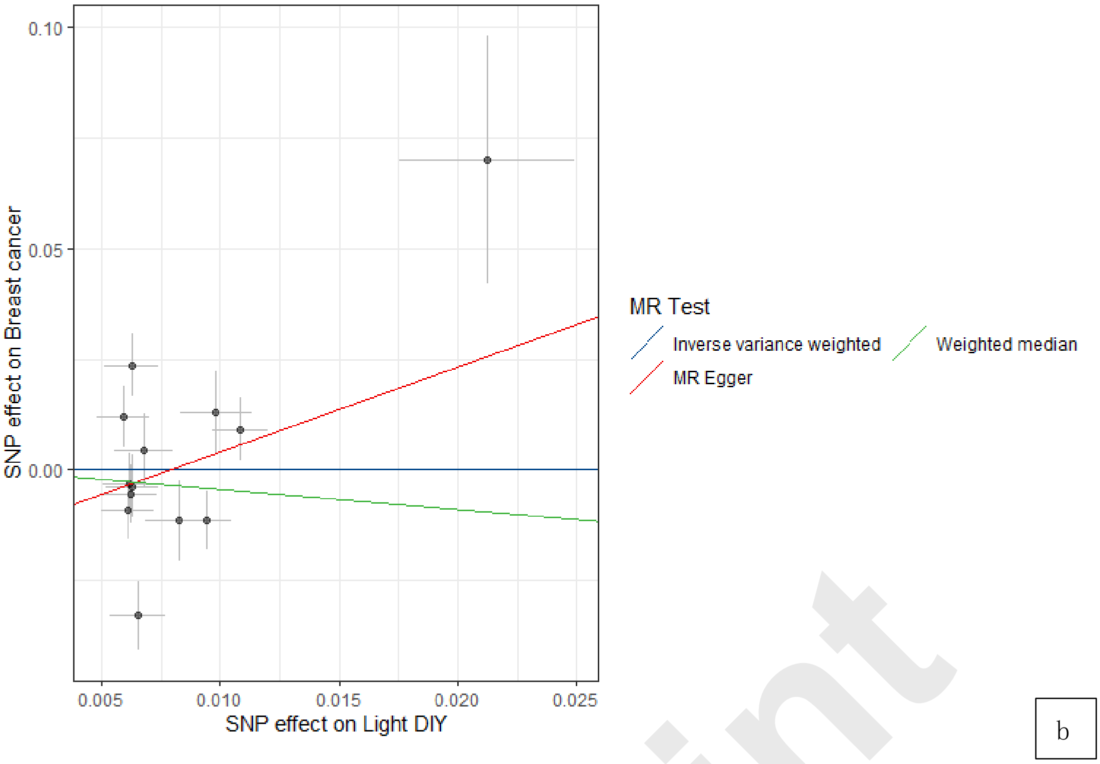
**Table 2.** MR analysis outcomes from several methods investigating the casual impact of distinct physical activities on breast cancer

| Exposure | nsnp | *SE* | IVW | | MR-Egger | | WMM | |
|---|---|---|---|---|---|---|---|---|
| | | | **OR/ß (95%*CI*)** | **p value** | **OR/ß (95%*CI*)** | **p value** | **OR/ß (95%*CI*)** | **p value** |
| **Breast cancer as outcome** | | | | | | | | |
| **H**eavy DIY | 18 | 0.421 | 0.739(0.324-1.689) | 0.474 | 0.139(0.001-254.7) | 0.614 | 0.880(0.394-1.963) | 0.755 |
| **L**ight DIY | 13 | 0.563 | 1.006(0.334-3.031) | 0.992 | 6.827(0.088-529.8) | 0.991 | 0.634(0.257-1.567) | 0.323 |
| **S**trenuous sports | 6 | 1.113 | 0.292(0.032-2.592) | 0.269 | 0.003(0.001-122.0) | 0.344 | 0.097(0.017- 0.541) | **0.008** |
| **W**alking for pleasure | 20 | 0.560 | 0.410(0.137-1.230) | 0.111 | 1.949(0.001-158978.8) | 0.909 | 0.638(0.256-1.591) | 0.909 |
| **O**ther exercises | 14 | 0.596 | 0.361(0.112-0.112) | 0.087 | 32.95(0.005-215785.3) | 0.450 | 0.580(0.255-1.318) | 0.193 |
| **ER+ breast cancer as outcome** | | | | | | | | |
| **H**eavy DIY | 18 | 0.420 | 0.529(0.232-1.206) | 0.130 | 0.081(0.001-149.4) | 0.522 | 0.431(0.182-1.022) | 0.056 |
| **L**ight DIY | 13 | 0.486 | 1.176(0.453-0.453) | 0.738 | 2.109(0.044-99.82) | 0.711 | 1.143(0.399-3.276) | 0.802 |
| **S**trenuous sports | 6 | 1.177 | 0.314(0.031-3.163) | 1.177 | 0.001(0.001-36.80) | 0.264 | 0.190(0.023-0.023) | 0.121 |
| **W**alking for pleasure | 20 | 0.541 | 0.302(0.105-0.872) | **0.027** | 2.602(0.001-142188.6) | 0.866 | 0.458(0.179-1.169) | 0.103 |
| **O**ther exercises | 14 | 0.605 | 0.510(0.156-1.672) | 0.267 | 24.27(0.002-199691.2) | 0.267 | 0.705(0.284-1.749) | 0.451 |

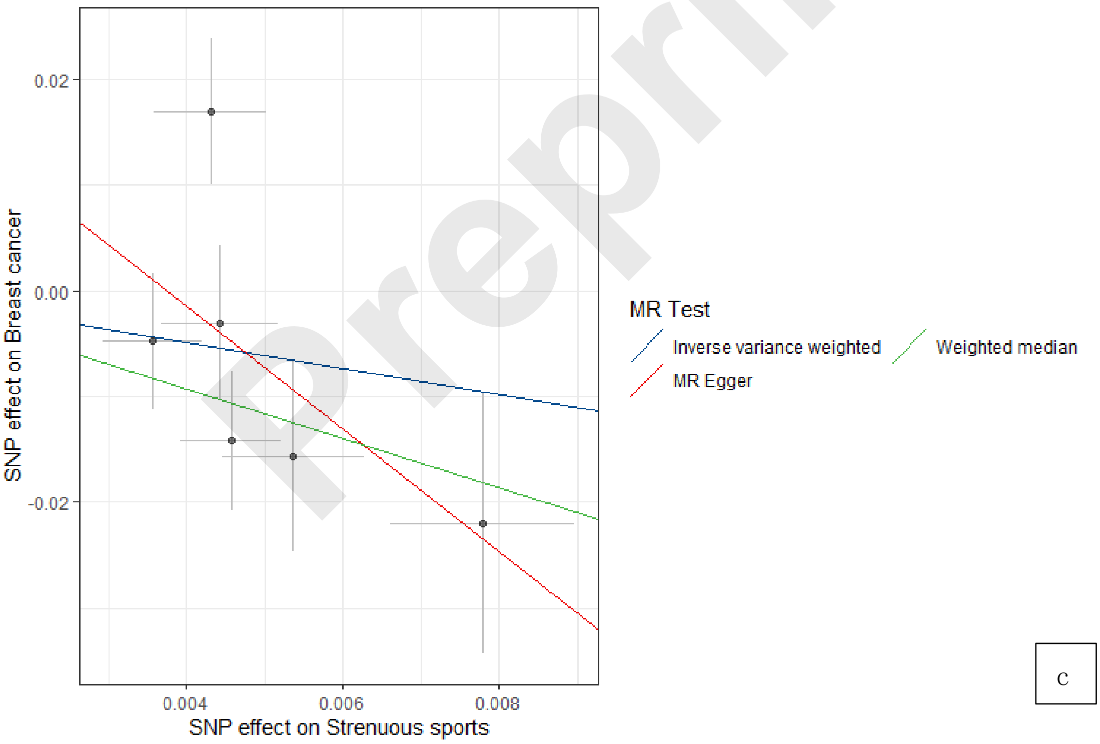
| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **ER- breast cancer as outcome** | | | | | | | | |
| Heavy DIY | 18 | 0.602 | 1.410(0.433-4.590) | 0.567 | 0.789(0.001-38910.5) | 0.966 | 0.689(0.166-2.850) | 0.607 |
| Light DIY | 13 | 1.092 | 0.571(0.067-4.858) | 0.608 | 61.63(0.014-261520.3) | 0.354 | 0.576(0.126-2.634) | 0.477 |
| Strenuous sports | 6 | 1.963 | 0.146(0.003-6.900) | 0.328 | 0.013(0.001-7032887) | 0.695 | 0.605(0.024-15.04) | 0.759 |
| Walking for pleasure | 20 | 0.585 | 0.477(0.151-1.504) | 0.207 | 87.05(0.001-9758134) | 0.462 | 0.541( 0.129-2.266) | 0.400 |
| Other exercises | 14 | 0.789 | 0.342(0.072-1.610) | 0.175 | 16.86(0.001-2573806) | 0.175 | 0.377(0.083-1.695) | 0.203 |

IVW = inverse variance weighted; WMM: weighed median method; SE: Standard Error; Breast cancer = overall breast cancer risk; ER = estrogen receptor-positive breast cancer risk; ER+− = estrogen receptor-negative breast cancer risk; Heavy DIY types of physical activity in last 4weeks: Heavy DIY (e.g., weeding, lawn mowing, carpentry, digging), Light DIY types of physical activity in last 4weeks: Light DIY (e.g., pruning, watering the lawn), Strenuous sports types of physical activity in last 4weeks: Strenuous sports, Walking for pleasure types of physical activity in last 4weeks: Walking for pleasure (not as a means of transport), Other exercises types of physical activity in last 4weeks: Other exercises (e.g., swimming, cycling, keep fit, bowling)




0.10
0.05
0.00
SNP effect on Breast cancer
0.005
0.010
0.015
0.020
0.025
SNP effect on Light DIY
MR Test
Inverse variance weighted
MR Egger
Weighted median

b


0.02
0.00
-0.02
SNP effect on Breast cancer
0.004
0.006
0.008
SNP effect on Strenuous sports
MR Test
Inverse variance weighted
MR Egger
Weighted median

c

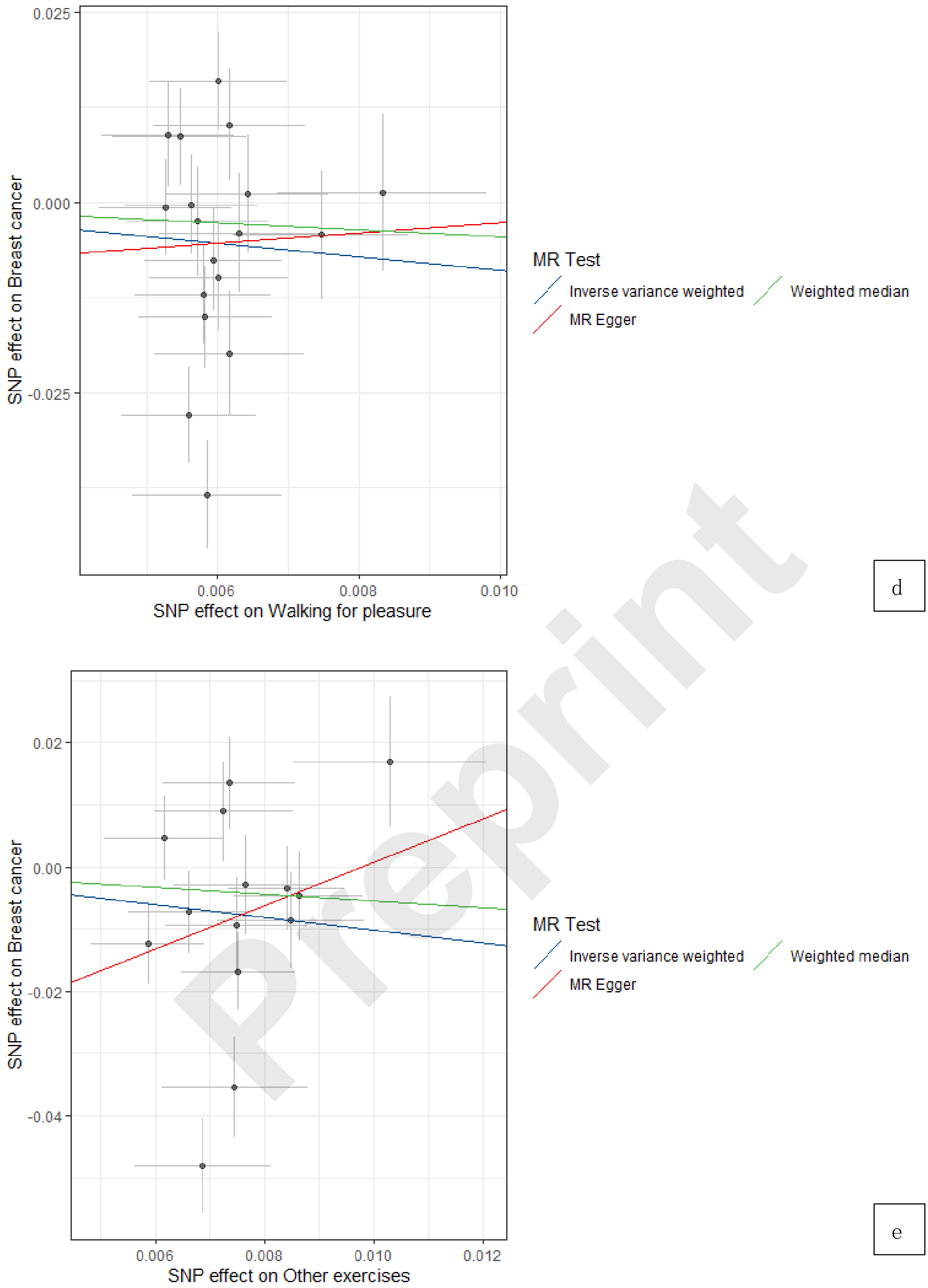
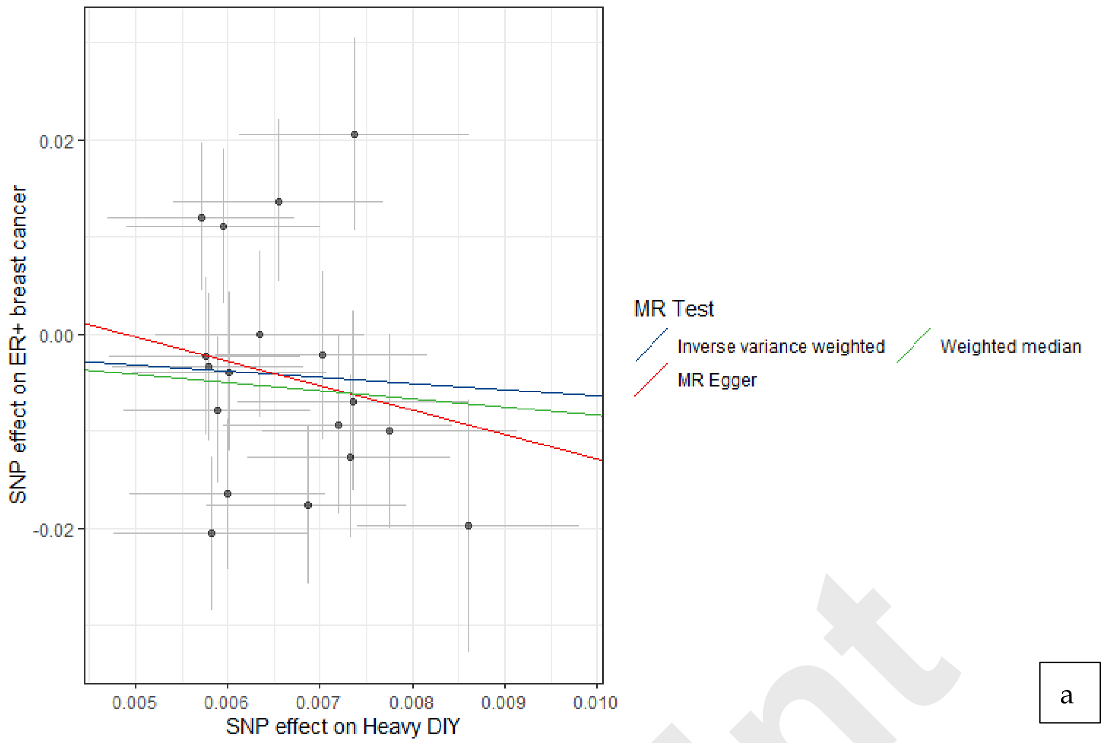
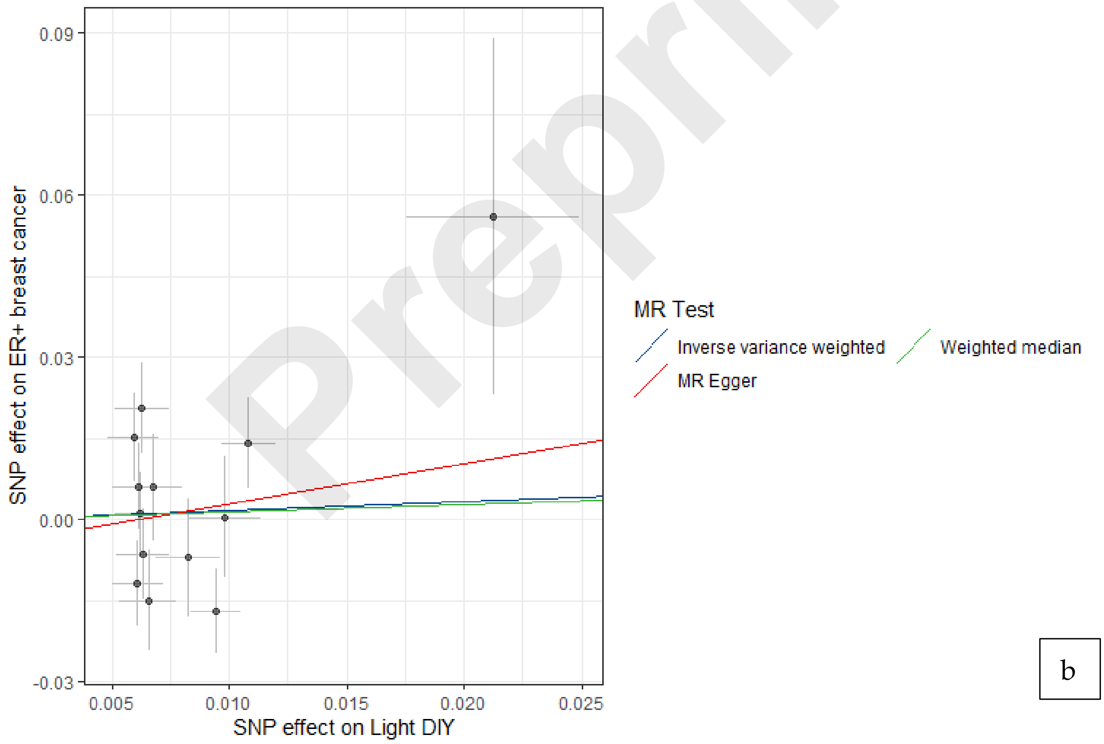



**Figure 2.** Scatter plots depicting the MR analysis of physical activities on BC for investigating casual impacts. a. Heavy DIY b. light DIY c. Strenuous sports d. Walking for pleasure e. Other exercises.



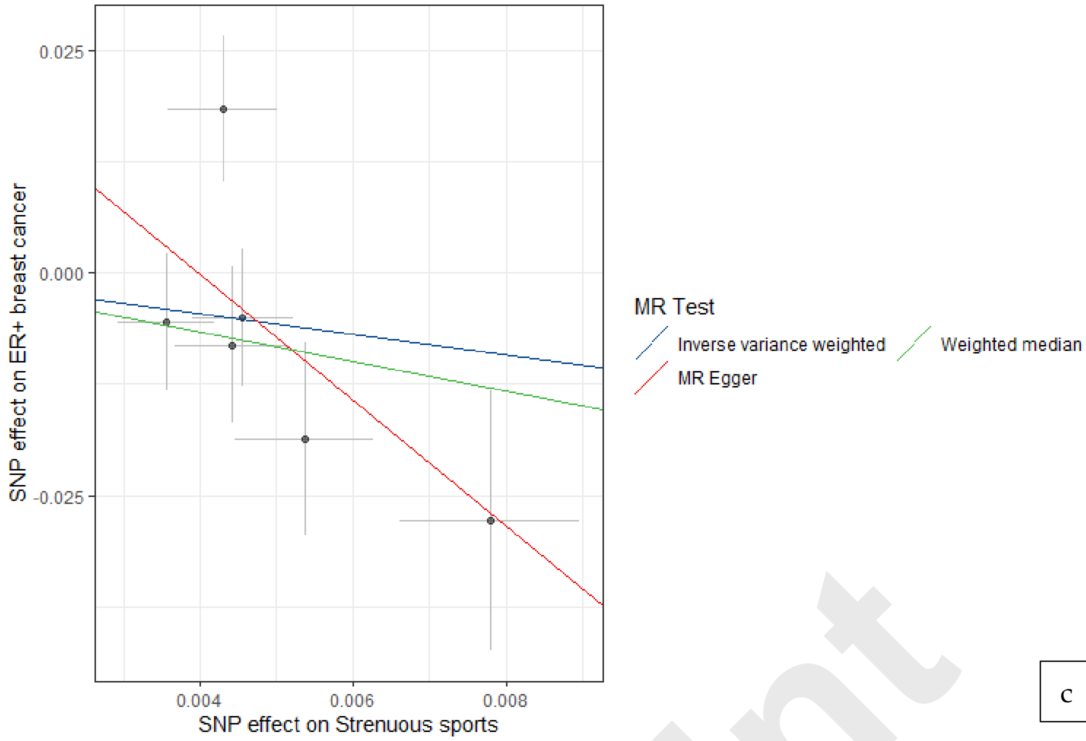
SNP effect on ER+ breast cancer
0.02
0.00
-0.02
0.005
0.006
0.007
0.008
0.009
0.010
SNP effect on Heavy DIY
MR Test
Inverse variance weighted
Weighted median
MR Egger


a

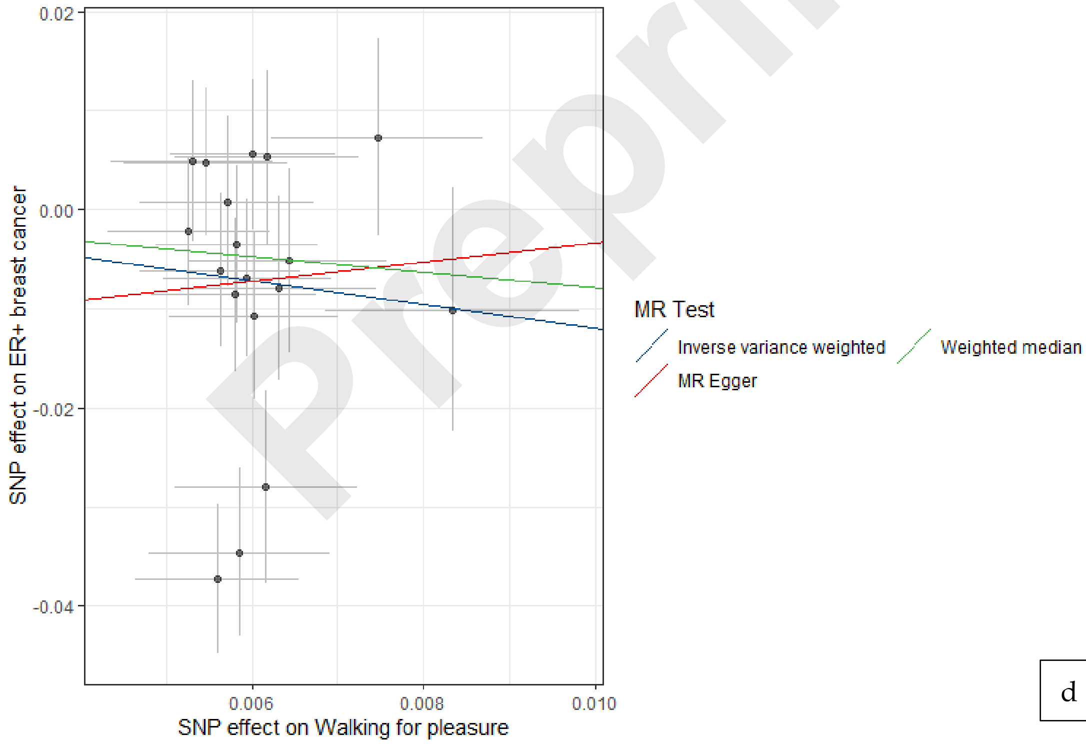


SNP effect on ER+ breast cancer
0.09
0.06
0.03
0.00
-0.03
0.005
0.010
0.015
0.020
0.025
SNP effect on Light DIY
MR Test
Inverse variance weighted
Weighted median
MR Egger

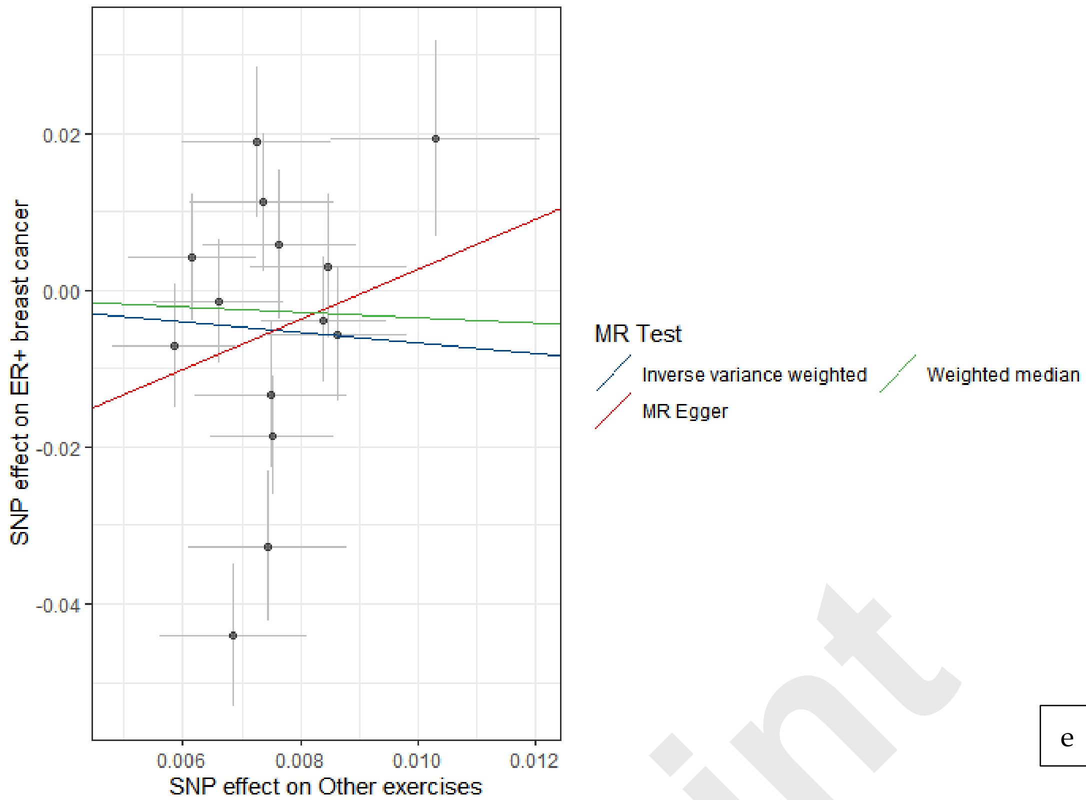

b


0.025
0.000
-0.025
SNP effect on ER+ breast cancer
0.004
0.006
0.008
SNP effect on Strenuous sports
MR Test
Inverse variance weighted
MR Egger
Weighted median

c


0.02
0.00
-0.02
-0.04
SNP effect on ER+ breast cancer
0.006
0.008
0.010
SNP effect on Walking for pleasure
MR Test
Inverse variance weighted
MR Egger
Weighted median

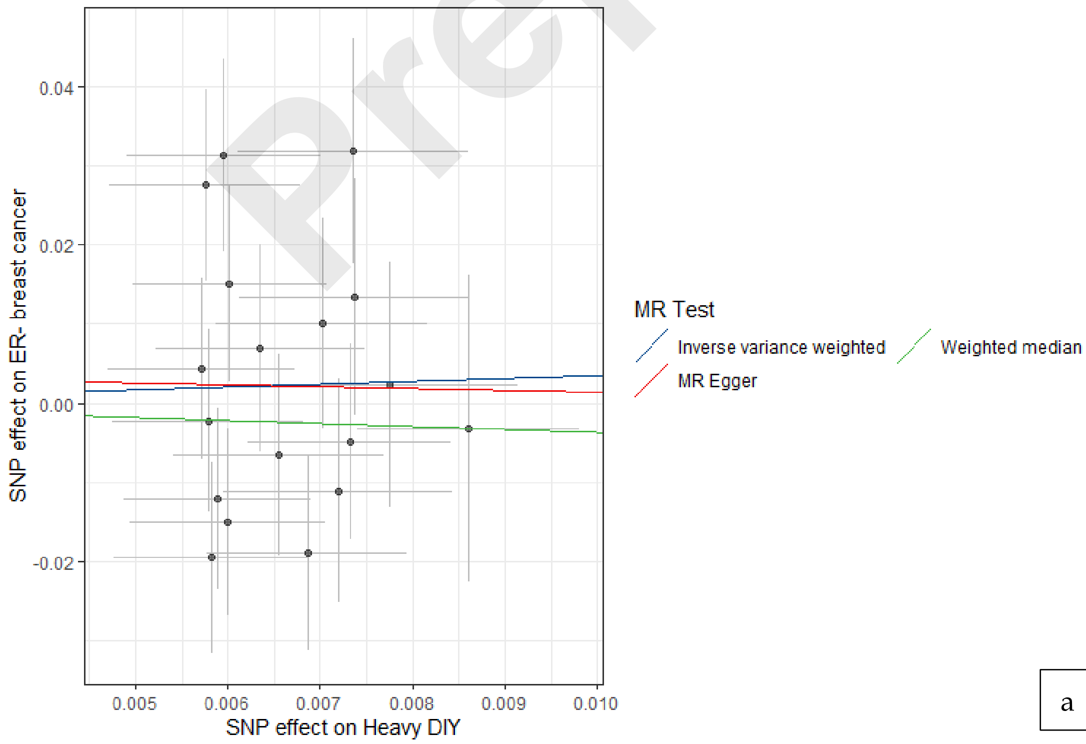
d



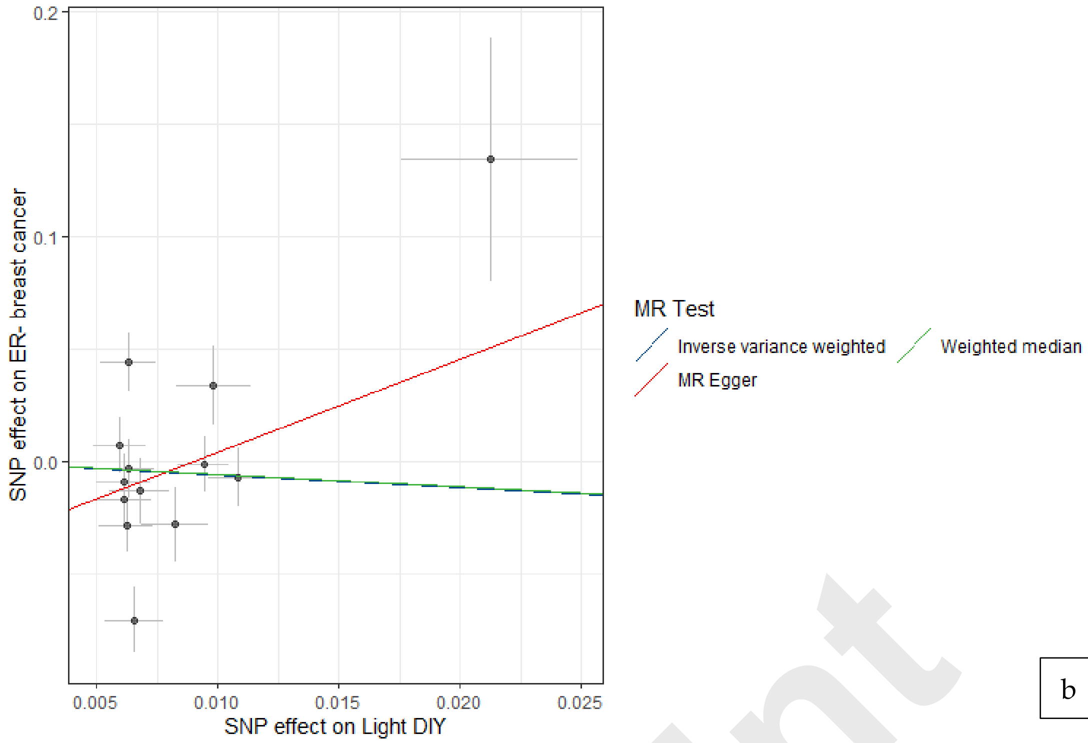

e

**Figure 3.** Scatter plots depicting the MR analysis of physical activities on ER+ BC for investigating casual impacts. a. Heavy DIY b. light DIY c. Strenuous sports d. Walking for pleasure e. Other exercises.

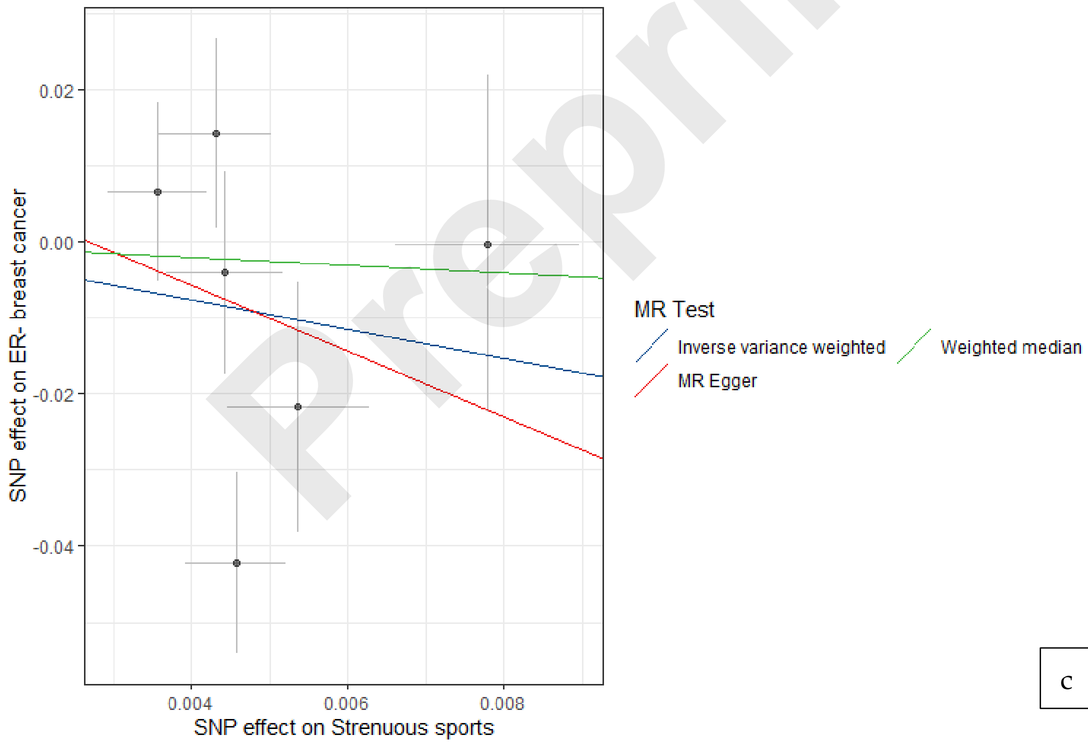





a


SNP effect on ER- breast cancer
0.2
0.1
0.0
0.005
0.010
0.015
0.020
0.025
SNP effect on Light DIY
MR Test
Inverse variance weighted
Weighted median
MR Egger

b


SNP effect on ER- breast cancer
0.02
0.00
-0.02
-0.04
0.004
0.006
0.008
SNP effect on Strenuous sports
MR Test
Inverse variance weighted
Weighted median
MR Egger

c

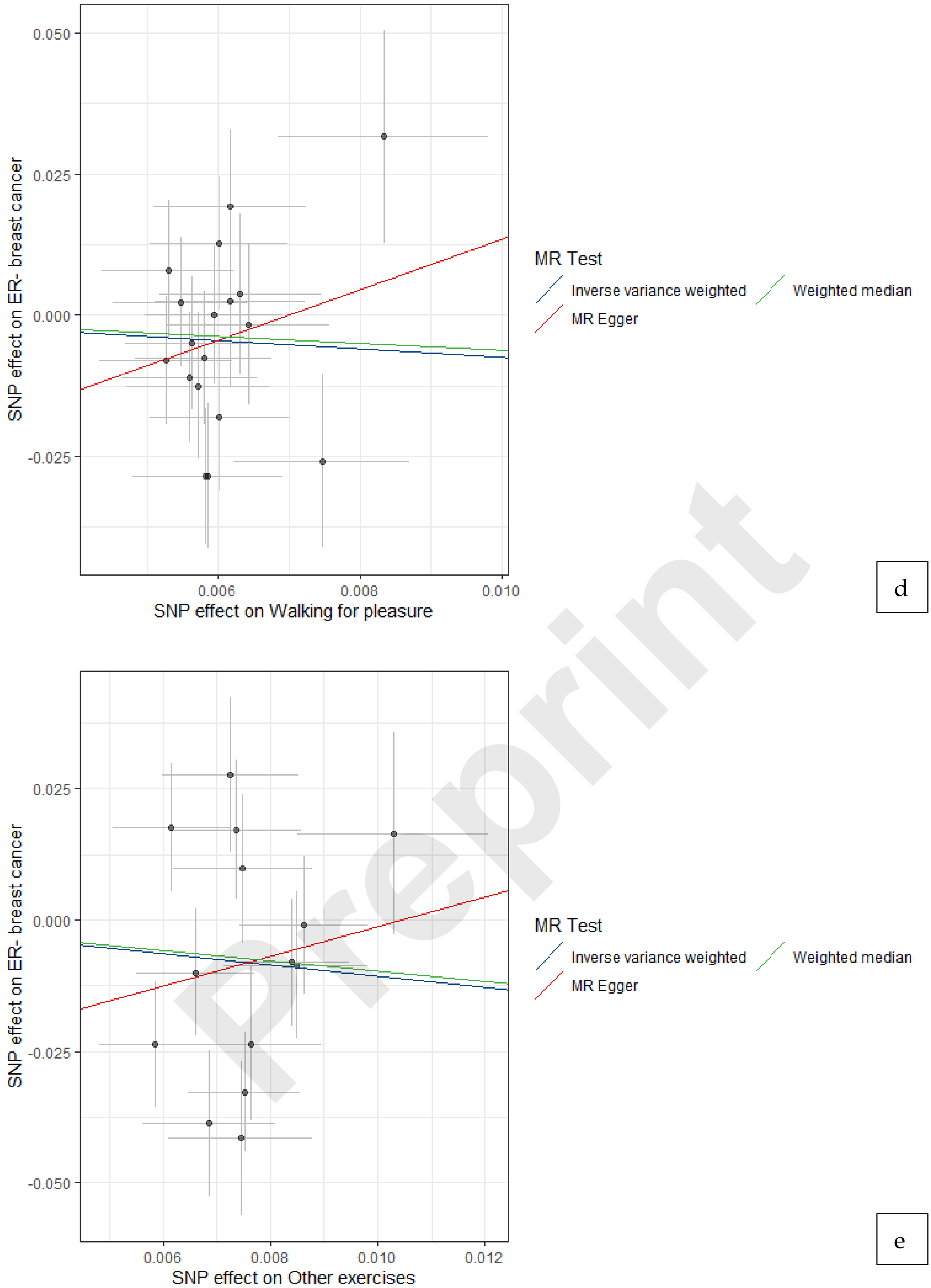





**Figure 4**. Scatter plots depicting the MR analysis of physical activities on ER-BC for investigating casual impacts. a. Heavy DIY b. light DIY c. Strenuous sports d. Walking for pleasure e. Other exercises.

### 3.2 Reverse MR analysis

BC, ER+ BC and ER- BC were utilized as exposures, while physical activities were employed as outcomes for conducting the reverse analysis. According to the estimates derived from the reverse MR analysis, no reverse causal association was observed between physical activities and breast cancer.

**Discussion**

With economic development and technological progress, there has been a gradual reduction in occupational, transportation, and daily physical activities, leading to a global issue of insufficient physical activity and increased sedentary behavior. This phenomenon has emerged as one of the most significant public health concerns of the 21st century. Research indicates that reduced physical activity is a crucial risk factor for cancer in women, including malignant tumors such as breast cancer, endometrial cancer, ovarian cancer, cervical cancer, and fallopian tube tumors.

Clinical studies have investigated the correlation between physical activities (PA) and breast cancer (BC). Prior research examining the potential impact of domain-specific physical activities on BC risk exhibits considerable heterogeneity[30]. On the one hand, BC demonstrate substantial cellular, genetic, and molecular heterogeneity[31]. The diversity forms the basis for the current clinical classification reliant on estrogen and progesterone receptors expression (ER and PR), as well as human epidermal growth factor receptor (HER2/ERBB2), facilitating targeted therapeutic approaches[32]. Conversely, physical activities encompass a multitude of daily movement patterns varying in timing, setting, and intensity, potentially leading to varied effects depending on type, intensity, and duration.[33]. Previous meta-analyses of prospective studies have suggested reduced BC risks associated with elevated PA levels. For instance, a case-control study in Addis Ababa shows significantly lower odds of BC among women engaging in moderate physical activities[34].However, another study in MCC-Spain reported that elevated levels of Moderate-to-high-intensity household (HPA) and recreational PA (RPA) correlated with decreased BC risk, exhibiting heterogeneity by molecular type, while sitting time consistently emerged as an independent risk factor for BC. The positive correlation observed between OPA (Occupational PA) and ER+/PR+ BC warrants further exploration[35]. Nevertheless, this review faces several limitations, including heterogeneity in the prescription of PA interventions (modality, frequency, duration, intensity, and timing), preclinical cancer models, and characteristics of human participants. However, it is important to acknowledge that this review encounters several constraints, notably the heterogeneity observed in the prescription of PA interventions, which encompasses variations in modality, frequency, duration, intensity, and timing. Additionally, the study is limited by the use of preclinical cancer models and the diverse characteristics of human participants. Moreover, a substantial proportion of the clinical studies analyzed are exploratory in nature, featuring small sample sizes, thereby hindering the formulation of definitive conclusions regarding the potential impact of PA on BC immune outcomes.

Several biological mechanisms have been postulated to elucidate the potential beneficial impacts of PA on BC progression. PA has been shown to decrease the concentrations of circulating insulin and insulin-like growth factor, stimulate cellular proliferation within breast tissues, and thereby inhibit cancer development in these tissues. Moreover, heightened levels of PA result in reduced circulating estradiol levels and elevated sex hormone-binding globulin levels, both of which are recognized risk factors for BC. Notably, the significant associations observed pertain to ER (estrogen receptor)-positive cancers rather than ER cancers alone, indicating that non-hormonal mechanisms may contribute to the protective effects of PA. This is the rationale behind our decision to stratify BC based on ER+ and ER- status and to investigate the causal relationship between PA and these subtypes.

Several biologic mechanisms have been suggested to elucidate the relationship between PA and BC risk[36-38]. The summary can be drawn as follows: Endogenous estrogen exposure, obesity, insulin-like growth factor I (IGF-I), and immune function[39]. 1) Endogenous estrogen exposure. Regular PA can reduce the occurrence of BC in women by reducing the accumulation and circulation of endogenous estrogen through late menarche age, early menopause, reduced frequency of menstruation, decreased estrogen levels in the follicular phase and progesterone levels in the luteal phase, as well as through the steroid hormone pathway. There is abundant epidemiological and clinical evidence demonstrating that estrogen significantly promotes the late-stage growth of estrogen-sensitive tumors, activating estrogen receptors to promote the proliferation of BC cells. Menopausal women with higher levels of daily PA have lower levels of estrogen precursors and higher levels of sex hormone-binding globulin[40, 41]. 2) Obesity. PA can reduce energy intake, leading to reduced postmenopausal weight, controlled weight gain, and decreased abdominal fat. Plasma levels of free estrogen increase in obese women, and after menopause, most of the estrogen in the blood comes from fat. Pre-menopausal obese women often experience cessation of ovulation, resulting in lower levels of circulating estrogen and progesterone. Therefore, PA reduces the risk of BC in postmenopausal women more significantly than in premenopausal women. 3) Insulin-like growth factor I (IGF-I). There is substantial evidence suggesting that androgens may increase the risk of ovarian cancer and BC, while progesterone has a protective effect[42-44]. IGF-I is a peptide hormone with functions and structures similar to insulin, stimulating all growth processes. Increased circulating concentrations of IGF-I can increase the risk of many cancers, such as BC[45]. Insulin indirectly increases the levels of biologically available estrogen and androgens by down-regulating sex hormone-binding globulin and up-regulating ovarian estrogen production, thereby increasing the risk of BC. A possible mechanism is that PA may lower serum levels of IGF-I by adjusting energy balance[46]. 4) Immune function. Many diseases are related to the body's immune function, and moderate PA can increase natural killer cells, lymphocytes, macrophages, and monocytes, thereby enhancing immune function. However, excessive PA may actually decrease immune function[47, 48].

The limitations of our MR study necessitate careful consideration. Our analysis was exclusively reliant on GWAS summary statistics derived from European populations, which inevitably confines the generalizability of our findings to other ethnic groups, notably Asians. Furthermore, the inability to compute sample overlap within this study is another constraint; however, the utilization of robust instrumental variables served as a partial mitigation for this potential bias. The exclusion of potential confounding factors, encompassing environmental determinants, occupational impacts, and BC treatments, posed a considerable challenge. Despite conducting an extensive array of sensitivity tests tailored to detect horizontal pleiotropy, the complete elimination of pleiotropic mechanisms remains impractical in the absence of comprehensive functional validations of these genetic loci. This limitation is primarily attributed to our limited understanding of the biological activities associated with these SNPs.

While vertical pleiotropy, where a single exposure influences an outcome through intermediary variables along the same causal chain, can be managed with appropriate statistical adjustments, horizontal pleiotropy—where an exposure affects multiple outcomes through distinct causal pathways—poses a formidable obstacle to MR inference. Addressing this complexity necessitates further advancements in our biological comprehension of these SNPs and the development of more sophisticated analytical methodologies.

In conclusion, our study utilized two-sample Mendelian randomization to deduce a causal link between PA and BC, concluding that walking for pleasure reliably influences ER+ BC risk. Our findings may offer valuable insights for clinical decision-making, suggesting that walking for pleasure may contribute to mitigating BC risk. If walking for pleasure indeed reduces the incidence risk of ER+ BC, then promoting physical exercise would be beneficial, not only for the general population—where it can bring public health benefits in terms of enhancing productivity and reducing healthcare costs—but also for those at risk of developing ER+ BC.

## Conclusion

In this study, we derived a robust conclusion through the implementation of the MR method, indicating a credible causal link between walking for pleasure and ER+ BC. Our findings imply a potential protective relationship between walking for pleasure and ER+ BC. Overall, our research lends support to the notion that undertaking walking for pleasure serves as an effective preventive measure against ER+ BC.

## Abbreviations

MR Mendelian randomization

BC Breast cancer

PA Physical activity

GWAS Genome-wide association studies

SNPs Single nucleotide polymorphisms

IVW Inverse variance weighting

WM Inverse variance weighting

MR-PRESSO Mendelian randomization pleiotropy residual sum and outlier

IVs Instrumental variables

OR Odds ratio

95% CI 95% Confidence interval

## Author contributions

Lishan DING: Conceptualization, Methodology, Software, Writing Original draft preparation; Qingliang CHEN and Hao LIANG: Visualization, Investigation; Ming ZHENG, Meng SHEN and Zhaojun LI: Supervision, Writing- Reviewing and Editing.

## Funding

This study was supported by the Medical Science and Technology Project of Henan Province (LHGJ20190855), Zhengzhou Municipal Science and Technology Innovation Guidance and Planning Project in the Healthcare Sector (2024YLZDJH056).

## Data availability

The data used in this study are publicly available. The summary statistics for physical activity of various types were acquired and extracted from an online public database (IEU Open GWAS Project https://gwas.mrcieu.ac.uk/). The summary data for Breast Cancer was extracted from the Breast Cancer Association Consortium (https://www.dkfz.de/en/epidemiologie-krebserkrankungen/units/genepi/ge_pr03_BCAC.).

## Declarations

### Ethics approval and consent to participate

Ethical approval and consent to participate is waived by the Research Ethics Committee of Henan No.3 Provincial People's Hospital is not required.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.



## references

[1] Sung H, Ferlay J, Siegel R L, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries[J]. CA Cancer J Clin, 2021,71(3):209-249.

[2] Benitez F J, Morgan E, de Luna A A, et al. Global Stage Distribution of Breast Cancer at Diagnosis: A Systematic Review and Meta-Analysis[J]. JAMA Oncol, 2024,10(1):71-78.

[3] Iacoviello L, Bonaccio M, de Gaetano G, et al. Epidemiology of breast cancer, a paradigm of the "common soil" hypothesis[J]. Semin Cancer Biol, 2021,72:4-10.

[4] Ehsan A N, Wu C A, Minasian A, et al. Financial Toxicity Among Patients With Breast Cancer Worldwide: A Systematic Review and Meta-analysis[J]. JAMA Netw Open, 2023,6(2):e2255388.

[5] Terry M B, Colditz G A. Epidemiology and Risk Factors for Breast Cancer: 21st Century Advances, Gaps to Address through Interdisciplinary Science[J]. Cold Spring Harb Perspect Med, 2023,13(9).

[6] Wang H, MacInnis R J, Li S. Family history and breast cancer risk for Asian women: a systematic review and meta-analysis[J]. BMC Med, 2023,21(1):239.

[7] Wang Z, Emmerich A, Pillon N J, et al. Genome-wide association analyses of physical activity and sedentary behavior provide insights into underlying mechanisms and roles in disease prevention[J]. Nat Genet, 2022,54(9):1332-1344.

[8] D'Amore C, Saunders S, Bhatnagar N, et al. Determinants of physical activity in community-dwelling older adults: an umbrella review[J]. Int J Behav Nutr Phys Act, 2023,20(1):135.

[9] Mazzucca S, Arredondo E M, Hoelscher D M, et al. Expanding Implementation Research to Prevent Chronic Diseases in Community Settings[J]. Annu Rev Public Health, 2021,42:135-158.

[10] Mishima R S, Verdicchio C V, Noubiap J J, et al. Self-reported physical activity and atrial fibrillation risk: A systematic review and meta-analysis[J]. Heart Rhythm, 2021,18(4):520-528.

[11] Lee I M, Shiroma E J, Lobelo F, et al. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy[J]. Lancet, 2012,380(9838):219-229.

[12] Hassen H Y, Ndejjo R, Musinguzi G, et al. Effectiveness of community-based cardiovascular disease prevention interventions to improve physical activity: A

systematic review and meta-regression[J]. Prev Med, 2021,153:106797.

[13] Tomida S, Koyama T, Ozaki E, et al. Seven-plus hours of daily sedentary time and the subsequent risk of breast cancer: Japan Multi-Institutional Collaborative Cohort Study[J]. Cancer Sci, 2024,115(2):611-622.

[14] Wang Z, Albers F E, Wang S E, et al. Biased effects of pre-diagnostic physical activity on breast cancer survival: Systematic review and meta-analysis[J]. Cancer Epidemiol, 2024,89:102544.

[15] Zhang Y, Song M, Yang Z, et al. Healthy lifestyles, systemic inflammation and breast cancer risk: a mediation analysis[J]. BMC Cancer, 2024,24(1):208.

[16] Sanchez-Bayona R, Gardeazabal I, Romanos-Nanclares A, et al. Leisure-time physical activity, sedentary behavior, and risk of breast cancer: Results from the SUN ('Seguimiento Universidad De Navarra') project[J]. Prev Med, 2021,148:106535.

[17] Mussallem D. Lifestyle for breast cancer risk reduction[J]. Menopause, 2022,29(8):979-981.

[18] Weitzer J, Castano-Vinyals G, Aragones N, et al. Effect of time of day of recreational and household physical activity on prostate and breast cancer risk (MCC-Spain study)[J]. Int J Cancer, 2021,148(6):1360-1371.

[19] Karavasiloglou N, Pestoni G, Kuhn T, et al. Adherence to cancer prevention recommendations and risk of breast cancer in situ in the United Kingdom Biobank[J]. Int J Cancer, 2022,151(10):1674-1683.

[20] Emdin C A, Khera A V, Kathiresan S. Mendelian Randomization[J]. JAMA, 2017,318(19):1925-1926.

[21] Sanderson E, Glymour M M, Holmes M V, et al. Mendelian randomization[J]. Nat Rev Methods Primers, 2022,2.

[22] Wang C, Wu W, Yang H, et al. Mendelian randomization analyses for PCOS: evidence, opportunities, and challenges[J]. Trends Genet, 2022,38(5):468-482.

[23] Yuan S, Mason A M, Titova O E, et al. Morning chronotype and digestive tract cancers: Mendelian randomization study[J]. Int J Cancer, 2023,152(4):697-704.

[24] Hu X, Zhao J, Lin Z, et al. Mendelian randomization for causal inference

accounting for pleiotropy and sample structure using genome-wide summary statistics[J]. Proc Natl Acad Sci U S A, 2022,119(28):e2106858119.

[25] Burgess S, Small D S, Thompson S G. A review of instrumental variable estimators for Mendelian randomization[J]. Stat Methods Med Res, 2017,26(5):2333-2355.

[26] Burgess S, Thompson S G. Interpreting findings from Mendelian randomization using the MR-Egger method[J]. Eur J Epidemiol, 2017,32(5):377-389.

[27] Wu G L, Liu Y Z, Zhang L. Editorial: Mendelian Randomization: Approach and Applications[J]. Front Genet, 2021,12:752146.

[28] Richmond R C, Davey S G. Mendelian Randomization: Concepts and Scope[J]. Cold Spring Harb Perspect Med, 2022,12(1).

[29] Cohen J F, Chalumeau M, Cohen R, et al. Cochran's Q test was useful to assess heterogeneity in likelihood ratios in studies of diagnostic accuracy[J]. J Clin Epidemiol, 2015,68(3):299-306.

[30] Fentie H, Ntenda P, Tiruneh F N. Dietary pattern and other factors of breast cancer among women: a case control study in Northwest Ethiopia[J]. BMC Cancer, 2023,23(1):1050.

[31] Akram M, Iqbal M, Daniyal M, et al. Awareness and current knowledge of breast cancer[J]. Biol Res, 2017,50(1):33.

[32] Donahue P, MacKenzie A, Filipovic A, et al. Advances in the prevention and treatment of breast cancer-related lymphedema[J]. Breast Cancer Res Treat, 2023,200(1):1-14.

[33] Herrmann S D, Willis E A, Ainsworth B E. The 2024 Compendium of Physical Activities and its expansion[J]. J Sport Health Sci, 2024,13(1):1-2.

[34] Hassen F, Enquselassie F, Ali A, et al. Association of risk factors and breast cancer among women treated at Tikur Anbessa Specialized Hospital, Addis Ababa, Ethiopia: a case-control study[J]. BMJ Open, 2022,12(9):e060636.

[35] Huerta J M, Molina A J, Chirlaque M D, et al. Domain-specific patterns of physical activity and risk of breast cancer sub-types in the MCC-Spain study[J]. Breast Cancer Res Treat, 2019,177(3):749-760.

[36] Turbitt W J, Xu Y, Sosnoski D M, et al. Physical Activity Plus Energy Restriction Prevents 4T1.2 Mammary Tumor Progression, MDSC Accumulation, and an Immunosuppressive Tumor Microenvironment[J]. Cancer Prev Res (Phila), 2019,12(8):493-506.

[37] Garritson J, Krynski L, Haverbeck L, et al. Physical activity delays accumulation of immunosuppressive myeloid-derived suppressor cells[J]. PLoS One, 2020,15(6):e0234548.

[38] Xu Y, Rogers C J. Impact of physical activity and energy restriction on immune regulation of cancer[J]. Transl Cancer Res, 2020,9(9):5700-5731.

[39] Wang B, Yuan F. The association between interleukin-1beta gene polymorphisms and the risk of breast cancer: a systematic review and meta-analysis[J]. Arch Med Sci, 2022,18(1):1-10.

[40] Gao L, Qi X, Hu K, et al. Estrogen receptor beta promoter methylation: a potential indicator of malignant changes in breast cancer[J]. Arch Med Sci, 2016,12(1):129-136.

[41] Kasielska-Trojan A, Danilewicz M, Struzyna J, et al. The role of oestrogen and progesterone receptors in gigantomastia[J]. Arch Med Sci, 2022,18(4):1016-1020.

[42] Lukanova A, Kaaks R. Endogenous hormones and ovarian cancer: epidemiology and current hypotheses[J]. Cancer Epidemiol Biomarkers Prev, 2005,14(1):98-107.

[43] Chung W M, Chen L, Chang W C, et al. Androgen/Androgen Receptor Signaling in Ovarian Cancer: Molecular Regulation and Therapeutic Potentials[J]. Int J Mol Sci, 2021,22(14).

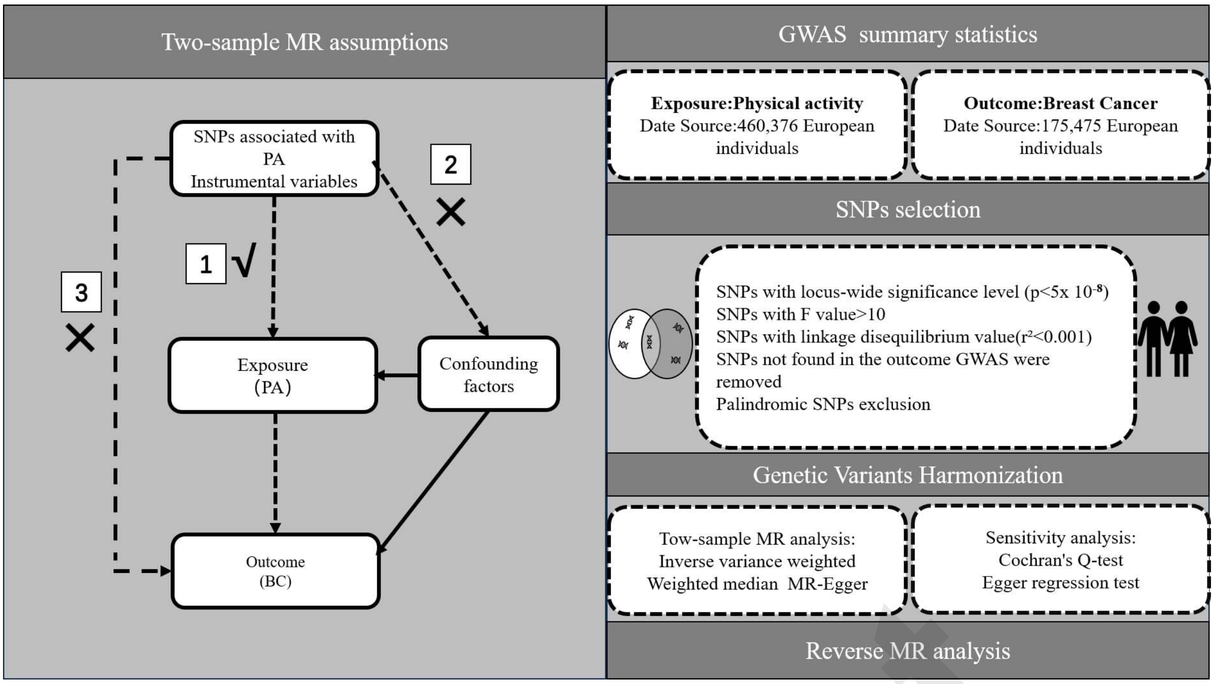
[44] Gjorgoska M, Rizner T L. The effect of androgens on the risk of endometriosis sub-phenotypes and ovarian neoplasms: A Mendelian randomization study[J]. J Steroid Biochem Mol Biol, 2024,239:106482.

[45] Voudouri K, Nikitovic D, Berdiaki A, et al. IGF-I/EGF and E2 signaling crosstalk through IGF-IR conduit point affects breast cancer cell adhesion[J]. Matrix Biol, 2016,56:95-113.

[46] Caffa I, Spagnolo V, Vernieri C, et al. Fasting-mimicking diet and hormone therapy induce breast cancer regression[J]. Nature, 2020,583(7817):620-624.

[47] Fang W, Zhou T, Shi H, et al. Progranulin induces immune escape in breast cancer via up-regulating PD-L1 expression on tumor-associated macrophages (TAMs) and promoting CD8(+) T cell exclusion[J]. J Exp Clin Cancer Res, 2021,40(1):4.

[48] Blomberg O S, Spagnuolo L, Garner H, et al. IL-5-producing CD4(+) T cells and eosinophils cooperate to enhance response to immune checkpoint blockade in breast cancer[J]. Cancer Cell, 2023,41(1):106-123.


Two-sample MR assumptions
SNPs associated with PA Instrumental variables
1 √
2 ×
3 ×
Exposure (PA)
Confounding factors
Outcome (BC)
GWAS summary statistics
Exposure:Physical activity
Date Source:460,376 European individuals
Outcome:Breast Cancer
Date Source:175,475 European individuals
SNPs selection
SNPs with locus-wide significance level (p<5x 10-8)
SNPs with F value>10
SNPs with linkage disequilibrium value(r²<0.001)
SNPs not found in the outcome GWAS were removed
Palindromic SNPs exclusion
Genetic Variants Harmonization
Tow-sample MR analysis:
Inverse variance weighted
Weighted median MR-Egger
Sensitivity analysis:
Cochran's Q-test
Egger regression test
Reverse MR analysis