# IEA

Mendelian Randomization

# Sex hormone binding globulin and risk of breast cancer: a Mendelian randomization study

Niki L Dimou,[1,2] Nikos Papadimitriou,[1,2] Dipender Gill,[3] Sofia Christakoudi,[3,4] Neil Murphy,[2] Marc J Gunter,[2] Ruth C Travis,[5] Tim J Key,[5] Renee T Fortner,[6] Philip C Haycock,[7,8] Sarah J Lewis,[7,8] Kenneth Muir,[9] Richard M Martin[7,8,10] and Konstantinos K Tsilidis[1,3]*

[1]Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece, [2]Section of Nutrition and Metabolism, International Agency for Research on Cancer, Lyon, France, [3]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK, [4]MRC Centre for Transplantation, King's College London, London, UK, [5]Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK, [6]Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany, [7]Bristol Medical School, Department of Population Health, University of Bristol, Bristol, UK, [8]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK, [9]Division of Population Health, Health Services Research and Primary Care, University of Manchester, Manchester, UK and [10]National Institute for Health Research (NIHR) Bristol Biomedical Research Centre, University Hospitals Bristol NHS Foundation Trust, and the University of Bristol, Bristol, UK

*Corresponding author. Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Stavros Niarchos Av., University Campus, Ioannina, Greece. E-mail: ktsilidi@cc.uoi.gr

## Abstract

**Background:** There are observational data suggesting an inverse association between circulating concentrations of sex hormone binding globulin (SHBG) and risk of postmenopausal breast cancer. However, causality is uncertain and few studies have investigated this association by tumour receptor status. We aimed to investigate these associations under the causal framework of Mendelian randomization (MR).

**Methods:** We used summary association estimates extracted from published genome-wide association study (GWAS) meta-analyses for SHBG and breast cancer, to perform two-sample MR analyses. Summary statistics were available for 122 977 overall breast cancer cases, of which 69 501 were estrogen receptor positive (ER$^{+ve}$) and 21 468 were ER$^{-ve}$, and 105 974 controls. To control for potential horizontal pleiotropy acting via body mass index (BMI), we performed multivariable inverse-variance weighted (IVW) MR as the main analysis, with the robustness of this approach further tested in sensitivity analyses.

**Results:** The multivariable IVW MR analysis indicated a lower risk of overall (odds ratio [OR]: 0.94; 95% confidence interval [CI]: 0.90, 0.98; *P*: 0.006) and ER$^{+ve}$ (OR: 0.92; 95% CI: 0.87, 0.97; *P*: 0.003) breast cancer, and a higher risk of ER$^{-ve}$ disease (OR: 1.09; 95%

CI: 1.00, 1.18; *P*: 0.047) per 25 nmol/L higher SHBG levels. Sensitivity analyses were consistent with the findings of the main analysis.

**Conclusions:** We corroborated the previous literature evidence coming from observational studies for a potentially causal inverse association between SHBG concentrations and risk of ER[+ve] breast cancer, but our findings also suggested a potential novel positive association with ER[-ve] disease that warrants further investigation, given the low prior probability of being true.

**Key words:** Sex hormone binding globulin, breast cancer, Mendelian randomization

**Key Messages**

- Using a Mendelian randomization analytical framework, we corroborated previous literature evidence coming from observational studies for a potentially causal inverse association between sex hormone binding globulin (SHBG) concentrations and risk of overall and estrogen receptor positive (ER[+ve]) breast cancer.
- Our findings also suggested a novel positive association with ER[-ve] disease, which warrants further investigation given the low prior probability of being true.
- Our study underlines that the role of SHBG in breast cancer development may be complex, potentially exerting differential effects depending on ER status.

## Introduction

Multiple lines of observational evidence suggest that endogenous sex steroid hormones play a central role in the development of breast cancer. Exposures related to elevated lifetime circulating estrogen concentrations, such as early age at menarche, nulliparity, late age at menopause and hormone replacement therapy, are well-established breast cancer risk factors.[1,2] Among postmenopausal women, positive associations of circulating estrogens and androgens with breast cancer are consistently reported in observational studies.[3] However, these associations are confirmed only for estrogen receptor-positive (ER[+ve]) breast cancer and the literature is sparse and inconsistent for estrogen receptor-negative (ER[-ve]) disease.[4–10] Sex hormone-binding globulin (SHBG) is a glycoprotein that binds sex steroid hormones. It plays a vital role in regulating concentrations of free estradiol and testosterone in circulation,[11] but may also have biological functions beyond sex hormone binding.[12,13] An inverse association between SHBG concentrations and risk of postmenopausal breast cancer has been consistently shown.[3,14] In contrast, associations by tumour receptor status have been inconsistent,[4,6,8,9] and for premenopausal disease they have been null.[15,16]

Residual confounding, reverse causation and exposure measurement error occur frequently in observational studies and may bias their results, hindering the ability to make robust causal inference. An alternative approach to conventional analyses of directly assessed exposures in observational studies is Mendelian randomization (MR). MR uses genetic variants robustly associated with the exposure of interest in an instrumental variable analysis, to make causal inferences about the effects of the exposure on an outcome.[17] The random and fixed allocation of alleles at conception makes confounding and reverse causation less likely explanations for associations identified in MR studies.[18]

Twin studies indicate that approximately half of the variance in circulating SHBG concentrations within populations is accounted for by genetic factors.[19] A meta-analysis of 10 genome-wide association studies (GWAS) in 21 791 individuals identified several genomic regions associated with circulating SHBG. These regions explained approximately 16% and 8% of the genetic variation in SHBG in men and women, respectively,[20] providing suitable genetic instruments to undertake MR analyses of genetically determined SHBG concentrations.

The aim of the present study was to investigate associations of genetically determined circulating SHBG concentrations with risk of overall breast cancer and risk stratified by ER status of the tumour under the MR causal framework. We used publicly available summary association data for 28 837 individuals with measured circulating SHBG concentrations and 122 977 breast cancer cases, adopting a two-sample MR design since the exposure and outcome were measured in separate non-overlapping samples.[21] To control for potential horizontal pleiotropy acting via body mass index (BMI) (Figure 1), we performed multivariable MR[22] because some of the selected genetic variants for SHBG were also associated with BMI.[23]

## Methods

### Data for the genetic epidemiology of SHBG and breast cancer

We selected genetic variants for the MR analysis on the basis of a genome-wide significant association with circulating SHBG concentrations (i.e. $P$-value threshold for inclusion at $< 5 \times 10^{-8}$). We extracted summary results for 13 variants reported in a GWAS meta-analysis of 28 837 people (13 899 women and 14 938 men) from 16 studies, which were adjusted for age, sex and BMI.[20] For these 13 variants (i.e. rs17496332, rs780093, rs3779195, rs440837, rs7910927, rs4149056, rs8023580, rs2411984, rs12150660, rs6258, rs1641537, rs1625895, rs1573036), we obtained publicly available summary association estimates for 122 977 women with overall breast cancer, of whom 69 501 women were $ER^{+ve}$ cases and 21 468 were $ER^{-ve}$ cases, and 105 974 controls. All women were of European ancestry, from the Breast Cancer Association Consortium (BCAC), and the GWAS analysis adjusted for principal components and country or study.[24] We excluded rs6258, having minor allele frequency <1% in the GWAS for breast cancer and exerting large effect size (Table 1).

### Statistical power

Power calculations were performed based on a method suggested by Brion et al.[26] We fixed the type-I error rate at 0.05. Under the current sample size, our study has 80% power to detect a causal effect of a relative 4% (i.e. OR: 0.96) decrease in breast cancer risk per 25 nmol/L higher SHBG concentrations, assuming an $R^2$ of 8% (variance explained by the selected SHBG variants). Corresponding estimates for $ER^{+ve}$ and $ER^{-ve}$ disease were 5% and 7% relative reductions, respectively. Assuming that a top-to-bottom quintile comparison is roughly equivalent to an OR per 2.8 standard deviation change (i.e. 25 nmol/L) in SHBG concentrations, our study had 80% power to detect ORs of 0.89, 0.87 and 0.82 or less comparing the top vs bottom quintiles of SHBG concentrations for overall, $ER^{+ve}$ and $ER^{-ve}$ breast cancer, respectively, which are smaller than the effect sizes observed in observational studies.[3,14] For completeness, we depict power calculations for a range of proportions of SHBG variation explained (Table 2).

### Statistical analysis

We employed a multivariable inverse-variance weighted (IVW) MR approach[22] to adjust for potential horizontal pleiotropy acting through BMI (Figure 1), because some of
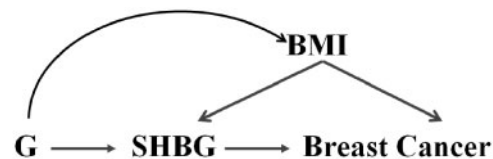


**Figure 1.** Graphical diagram of the Mendelian randomization analysis between sex hormone binding globulin (SHBG) concentrations and risk of breast cancer. BMI, body mass index; G, gene or genetic instrument.

the selected genetic variants for SHBG (i.e. rs12150660, rs1625895, rs7910927, rs780093 and rs17496332) were also associated with BMI (smallest $P$-value $7.64 \times 10^{-5}$ for rs780093),[23] and BMI has been consistently associated with SHBG concentrations[27] and breast cancer risk.[28,29] Publicly available genetic data for BMI were retrieved from the GIANT consortium for 339 000 individuals, 95% of whom were of European descent[23] (Table 1). We also applied the multivariable MR-Egger method to investigate for potential pleiotropic pathways other than via BMI.[30] For comparison, we employed two univariable MR methods, a fixed-effects IVW average of single nucleotide polymorphism (SNP)-specific associations and a likelihood-based method,[31,32] which do not take into account potential horizontal pleiotropy. For ease of comparison with observational studies, we transformed beta coefficients from the logarithmic scale, which were originally reported in the published GWAS,[20] into the natural scale using a formula suggested by Rodriguez-Barranco et al.[25] All MR effect estimates are reported as odds ratios (OR) per standard deviation (i.e. 25 nmol/L) higher SHBG concentrations.

A series of statistical tests were performed to investigate the potential violation of MR assumptions.[33,34] The first assumption (i.e. that the genetic variants are strongly associated with circulating SHBG concentrations) was very likely satisfied by employing genetic variants associated with circulating SHBG concentrations at a genome-wide significance level. To test for potential violation of the second and third MR assumptions (i.e. that the genetic variants are not associated with any confounder of the SHBG-breast cancer association and are conditionally independent of breast cancer, given SHBG concentrations and all confounders), we acquired information for the association of the selected SHBG SNPs with other traits from the GWAS Catalogue.[35] To further statistically probe for existence of horizontal pleiotropy, which means that the selected variants have an effect on other traits outside the pathway of SHBG and have an impact on breast cancer risk violating the third MR assumption, we employed the Cochran's Q statistic that quantifies the heterogeneity in effect sizes attributed to the selected genetic variants.

**Table 1.** Characteristics of genetic variants associated with sex hormone-inding globulin (SHBG) and breast cancer in published GWAS

| SNPs | Chr: pos (hg19) | Locus | Effect/ref. allele | Circulating SHBG | | Body mass index | | Overall breast cancer | | ER^+ve breast cancer | | ER^-ve breast cancer | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Beta$^a$ (SE) | P-value | Beta$^b$ (SE) | P-value | Beta$^c$ (SE) | P-value | Beta$^c$ (SE) | P-value | Beta$^c$ (SE) | P-value |
| rs17496332 | 1: 107546375 | PRMT6 | G/A | 0.028 (0.0041) | 1.40E-11 | 0.0124 (0.0032) | 1.01E-04 | 0.0003 (0.0068) | 9.61E-01 | -0.0009 (0.0081) | 9.13E-01 | 0.0169 (0.0123) | 1.68E-01 |
| rs780093 | 2: 27742603 | GCKR | C/T | 0.032 (0.0039) | 2.20E-16 | 0.012 (0.0030) | 7.64E-05 | 0.0113 (0.0064) | 7.55E-02 | 0.0106 (0.0076) | 1.61E-01 | 0.0136 (0.0115) | 2.39E-01 |
| rs3779195 | 7: 97993362 | BAIAP2L1 | T/A | 0.028 (0.0051) | 2.70E-08 | -0.001 (0.0047) | 8.32E-01 | -0.0040 (0.0081) | 6.23E-01 | -0.0072 (0.0096) | 4.50E-01 | 0.0253 (0.0148) | 8.76E-02 |
| rs440837 | 8: 81461974 | ZBTB10 | G/A | 0.028 (0.0047) | 3.40E-09 | -0.0058 (0.0045) | 1.97E-01 | -0.0088 (0.0075) | 2.46E-01 | -0.0090 (0.0090) | 3.18E-01 | 0.0183 (0.0136) | 1.78E-01 |
| rs7910927 | 10: 65138910 | JMJD1C | G/T | 0.048 (0.0039) | 6.10E-35 | -0.0124 (0.0036) | 5.72E-04 | -0.0217 (0.0062) | 5.24E-04 | -0.0286 (0.0075) | 1.32E-04 | -0.0018 (0.0115) | 8.73E-01 |
| rs4149056 | 12: 21331549 | SLCO1B1 | T/C | 0.029 (0.0052) | 1.90E-08 | -0.0002 (0.004) | 9.51E-01 | -0.0098 (0.0084) | 2.45E-01 | -0.0115 (0.0100) | 2.49E-01 | -0.0130 (0.0153) | 3.93E-01 |
| rs8023580 | 15: 96708291 | NR2F2 | C/T | 0.030 (0.0044) | 8.30E-12 | -0.0008 (0.0042) | 8.49E-01 | -0.0085 (0.0076) | 2.61E-01 | -0.0144 (0.0090) | 1.09E-01 | 0.0067 (0.0135) | 6.20E-01 |
| rs2411984 | 17: 47445751 | ZNF652 | A/G | 0.033 (0.0044) | 3.50E-14 | 0.0001 (0.004) | 9.80E-01 | -0.0078 (0.0068) | 2.49E-01 | -0.0009 (0.0081) | 9.11E-01 | 0.0044 (0.0123) | 7.23E-01 |
| rs12150660 | 17: 7521915 | SHBG | T/G | 0.103 (0.0047) | 1.80E-106 | 0.0091 (0.0043) | 3.43E-02 | -0.0043 (0.0072) | 5.51E-01 | -0.0018 (0.0086) | 8.35E-01 | 0.0068 (0.0131) | 6.05E-01 |
| rs1641537 | 17: 7545721 | SHBG | C/T | 0.0814 (0.0062) | 8.19E-39 | -0.0014 (0.0054) | 7.95E-01 | -0.0042 (0.0094) | 6.52E-01 | -0.0139 (0.0111) | 2.11E-01 | 0.0317 (0.0173) | 6.67E-02 |
| rs1625895 | 17: 7578115 | TP53 | C/T | 0.052 (0.0067) | 1.17E-14 | 0.0134 (0.0059) | 2.31E-02 | 0.0209 (0.0093) | 2.42E-02 | 0.0209 (0.0111) | 6.00E-02 | 0.0191 (0.0170) | 2.61E-01 |
| rs1573036 | X: 109820068 | TDGF3 | T/C | 0.028 (0.0037) | 4.10E-14 | NA | NA | -0.0053 (0.0081) | 5.16E-01 | -0.0002 (0.0093) | 9.81E-01 | -0.0066 (0.0145) | 6.49E-01 |

GWAS, genome-wide association studies; SNPs, single nucleotide polymorphisms; Chr, chromosome; pos, position; SHBG, sex hormone binding globulin; ER, estrogen receptor; SE, standard error; NA, not available.

$^a$Beta units are per-allele effect estimates in natural logarithm-transformed SHBG concentrations (nmol/L),[20] To enable better comparison with results from observational studies, we run MR analyses after transforming these beta coefficients into the natural scale (nmol/L) using a formula suggested by Rodriguez-Barranco and colleagues.[25]

$^b$Beta units are per standard deviation increase of body mass index (kg/m$^2$),[23]

$^c$Per-allele logarithm of the odds ratios between breast cancer cases and controls.[24]

**Table 2.** Number of cancer cases and controls and statistical power in Mendelian randomization study of SHBG and breast cancer risk

| Cancer type | Cases | Controls | Total | Proportion of cases | Minimum detectable odds ratio[a] | | |
|---|---|---|---|---|---|---|---|
| | | | | | $R^2 = 0.06$/F statistic = 153.3 | $R^2 = 0.08$/F statistic = 208.9 | $R^2 = 0.10$/F statistic = 266.9 |
| Overall | 122 977 | 105 974 | 228 951 | 0.54 | 0.953/1.049 | 0.960/1.042 | 0.963/1.038 |
| ER$^{+ve}$ | 69 501 | 95 042 | 164 543 | 0.42 | 0.945/1.058 | 0.952/1.050 | 0.957/1.045 |
| ER$^{-ve}$ | 21 468 | 100 594 | 122 062 | 0.18 | 0.922/1.085 | 0.930/1.075 | 0.937/1.067 |

[a]Minimum detectable odds ratio per 25 nmol/L increase/decrease in SHBG levels: assume 80% power, 5% alpha level and that 6% to 10% of SHBG variance is explained by the 12 SNPs used in the MR analysis.

When there was evidence for heterogeneity, we performed a random-effects IVW approach in order to take into account this source of uncertainty.[36] MR-Egger regression was also used, where values away from zero for the intercept term are an indication of horizontal pleiotropy.[37] The slope of the MR-Egger regression[37] and the estimator from the weighted median[38] and weighted mode[39] approaches were used to estimate causal effects accounting for potential violations of the second and third MR assumptions. The MR pleiotropy residual sum and outlier test (MR-PRESSO) was also used to identify pleiotropic variants (P-value threshold set at 0.05) and if there was evidence for pleiotropy, those variants were excluded.[40] The weighted median, weighted mode and MR-PRESSO analyses were performed only in the univariable MR approach, as these methods have not been extended in multivariable MR.

Further, sensitivity analyses were performed to test the robustness of the genetic instrument for SHBG concentrations after: (i) excluding one SNP (i.e. rs780093) due to potential pleiotropy with several other traits (e.g. urate levels, triglycerides, Crohn's disease, breast size[41–44]); (ii) excluding three SNPs (i.e. rs1641537, rs1625895 and rs3779195) that were derived from conditional analyses (i.e. adjusting for other genetic variants) in the GWAS for SHBG; (iii) using only three SNPs (i.e. rs12150660, rs7910927, rs780093) that were genome-wide significant in the GWAS analysis only among women; (iv) using female-specific estimates for the SNP-SHBG associations (for three SNPs, i.e. rs1641537, rs1625895 and rs3779195, estimates were only reported in males and females together); (v) using only two SNPs in the SHBG gene as instruments (i.e. rs12150660 and rs1641537); and (vi) excluding five SNPs (i.e. rs12150660, rs1625895, rs7910927, rs780093 and rs17496332) associated with BMI. Sensitivity analyses (v) and (vi) were performed only in the univariable MR framework.

All the statistical analyses were implemented in the Mendelian randomization R package,[45] apart from the weighted mode approach where we used the MR robust package in STATA.[46]

## Results

### Mendelian randomization estimates

Figure 2 shows the multivariable IVW MR analysis adjusting for the potential horizontal pleiotropy via BMI. A 6% decreased risk for overall breast cancer was observed per 25 nmol/L higher SHBG concentrations [odds ratio (OR): 0.94; 95% confidence interval (CI): 0.90, 0.98; P: 0.006] and an 8% decreased risk for ER$^{+ve}$ disease (OR: 0.92; 95% CI: 0.87, 0.97; P: 0.003). In contrast, there was a 9% increased risk for ER$^{-ve}$ disease per 25 nmol/L higher SHBG concentrations (OR: 1.09; 95% CI: 1.00, 1.18; P: 0.047). There was little evidence of heterogeneity in the effect sizes attributed to each of the genetic variants for associations with overall (Cochran's Q P: 0.74), ER$^{+ve}$ (P: 0.75) and ER$^{-ve}$ breast cancer (P: 0.55). The multivariable MR-Egger analysis yielded large P-values for the intercept term, indicating low probability of horizontal pleiotropy; the point estimates of the slope were consistent with our main MVMR IVW analysis, but the confidence intervals were wider for overall breast cancer (OR: 0.97; 95% CI: 0.88, 1.08; P: 0.572; P-intercept: 0.385) and by tumour receptor status (ER$^{+ve}$ OR: 0.97; 95% CI: 0.86, 1.10; P: 0.564; P-intercept: 0.325 and ER$^{-ve}$ disease OR: 1.00; 95% CI: 0.83, 1.21; P: 0.970; P-intercept: 0.294), but this method is known to have low power when few genetic instruments are used.[47]

When we performed the univariable IVW MR analysis that does not account for potential horizontal pleiotropic effects via BMI (Supplementary Table 1, available as Supplementary data at IJE online), the results were similar to the multivariable IVW analysis, but they were slightly attenuated for overall (OR: 0.96; 95% CI: 0.92, 1.00; P: 0.07) and ER$^{+ve}$ breast cancer (OR: 0.95; 95% CI: 0.91, 1.00; P: 0.06). The results were almost identical for ER$^{-ve}$
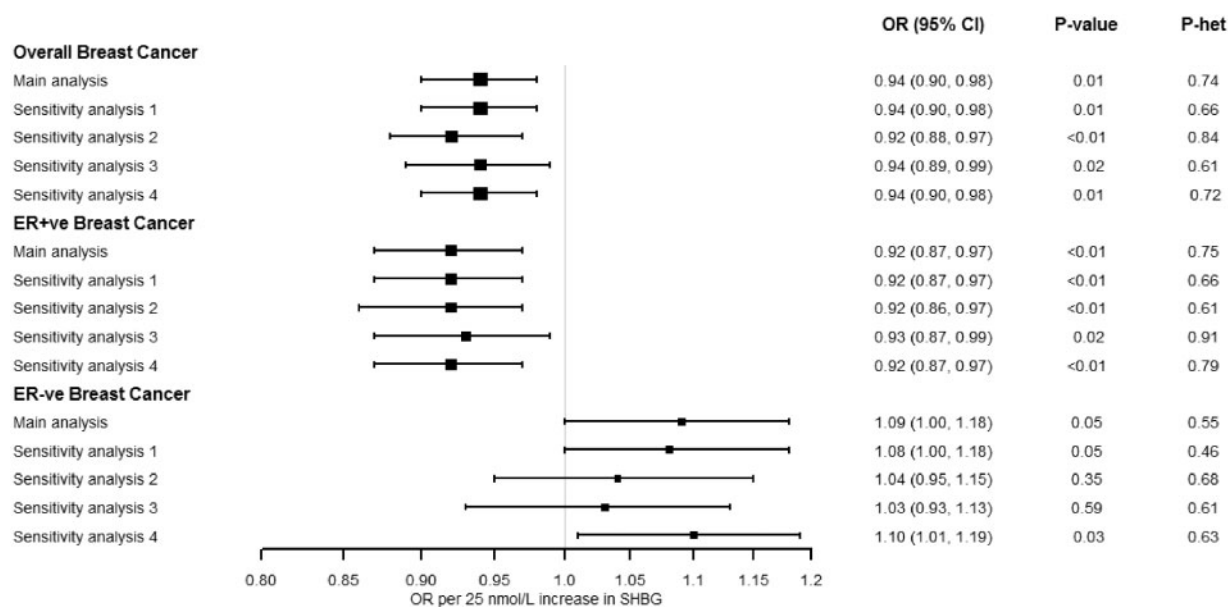
**Figure 2**. Multivariable inverse-variance weighted Mendelian randomization estimates between sex hormone binding globulin (SHBG) concentrations and risk of breast cancer, adjusting for the genetic effects of body mass index (BMI). Main analysis: the odds ratios represent increase/decrease of risk per 25nmol/L increase in SHBG levels ($N = 12$ SNPs). Sensitivity analysis 1: we used 11 SNPs after excluding rs780093 due to potential pleiotropy with several other traits.[41–44] Sensitivity analysis 2: we used nine SNPs after excluding rs1641537, rs1625895 and rs3779195 derived from conditional analyses in the GWAS of SHBG.[20] Sensitivity analysis 3: we used as instruments only the three SNPs (i.e. rs12150660, rs7910927, rs780093), which were significant in the GWAS analysis for SHBG only in women.[20] Sensitivity analysis 4: we used female-specific estimates for the SNP-SHBG associations (for three SNPs i.e. rs1641537, rs1625895 and rs3779195, estimates were only reported in males and females together).[20].

breast cancer (OR: 1.09; 95% CI: 1.01, 1.18; *P*: 0.03). The maximum likelihood univariable MR approach yielded almost identical results.

## Sensitivity analyses

The multivariable IVW results for overall and ER$^{+ve}$ breast cancer did not change in sensitivity analyses that removed genetic variants from the instrument of SHBG to test its robustness (Figure 2). The association for ER$^{-ve}$ disease remained after excluding rs780093 (sensitivity analysis 1) and when using female-specific SNP-SHBG association estimates (sensitivity analysis 4), but it was not observed in other sensitivity analyses (Figure 2).

We applied several statistical tests and sensitivity analyses in the univariable IVW MR approach to further test the robustness of MR assumptions (Supplementary Table 1). There was some evidence of heterogeneity for associations of SHBG with overall (Cochran's Q *P*: 0.01) and ER$^{+ve}$ (*P*: 0.02) breast cancer. However, the random-effects IVW analyses provided similar estimates with only slightly wider confidence intervals for overall (OR: 0.96; 95% CI: 0.90, 1.02) and ER$^{+ve}$ disease (OR: 0.95; 95% CI: 0.88 1.02). The MR-Egger intercept yielded large *P*-values. suggesting absence of horizontal pleiotropy, but this analysis was likely underpowered due to the relatively small number of genetic variants

(Supplementary Table 1). When we applied the MR-Egger regression slope approach, the weighted median and the weighted mode approaches, the point estimates were on the same direction as the IVW approach but the *P*-values were large (Supplementary Table 1). The MR-PRESSO test indicated one SNP, rs7910927, as an outlier for overall and ER$^{+ve}$ disease, which was also evident when we estimated and plotted MR results by each separate SNP (Supplementary Figures 1–3, available as Supplementary data at *IJE* online). When this variant was excluded from the multivariable IVW analysis, the results were very similar with the multivariable analysis including all SNPs (overall breast cancer, OR: 0.95; 95% CI: 0.90, 1.00; *P*: 0.045; ER$^{+ve}$, OR: 0.94; 95% CI: 0.88, 1.00; *P*: 0.046; ER$^{-ve}$, OR: 1.10; 95% CI: 1.00, 1.22; *P*: 0.051). In addition, when rs7910927 variant was excluded along with other four variants associated with BMI (sensitivity analysis 6) and univariable MR was run, we observed evidence for association for overall (OR: 0.93; 95% CI: 0.86, 1.00; *P*: 0.04), ER$^{+ve}$ (OR: 0.91; 95% CI: 0.83, 0.99; *P*: 0.03) and ER$^{-ve}$ breast cancer (OR: 1.15; 95% CI: 1.00, 1.32; *P*: 0.04) in agreement with the results from the multivariable IVW analysis. Similar evidence for association was observed in most other sensitivity analyses that removed genetic variants from the instrument of SHBG to test its robustness (Supplementary Table 1).

## Discussion

We conducted a large MR study using summary statistics based on 122 977 women with breast cancer, of whom 69 501 cases had ER$^{+ve}$ disease and 21 468 cases had ER$^{-ve}$ disease. We demonstrated for the first time under the MR causal framework an inverse association of genetically determined SHBG concentrations with risk of overall and ER$^{+ve}$ breast cancer, but a positive association for ER$^{-ve}$ disease.

A substantial number of observational studies have assessed the association of circulating SHBG concentrations with risk of postmenopausal breast cancer. A meta-analysis of 26 prospective studies involving 5172 postmenopausal breast cancer cases and 10 939 controls estimated an OR of 0.64 (95% CI; 0.57, 0.72) comparing the highest versus the lowest concentrations of SHBG, and had low between-study heterogeneity and little evidence of small-study effects.[14] Similar results were observed in a pooled analysis of nine prospective studies.[3] These findings are concordant with the results of the current MR study. Assuming that a top-to-bottom quintile comparison is roughly equivalent to an OR per 2.8 standard deviations (i.e. 25 nmol/L), our MR study estimated an OR equal to 0.84 (95% CI: 0.74, 0.95) for overall breast cancer risk, 0.80 (95% CI: 0.69, 0.93) for ER$^{+ve}$ and 1.26 (95% CI: 1.00, 1.58) for ER$^{-ve}$ breast cancer.

The literature on the association of circulating SHBG concentrations with breast cancer risk stratified by tumour receptor status is limited. The largest available study using data from 382 postmenopausal ER$^{+ve}$ (602 controls) and 172 ER$^{-ve}$ breast cancer cases (219 controls) suggested an inverse association for ER$^{+ve}$ disease (OR: 0.71; 95% CI: 0.51, 1.00) and a similar but imprecise association for ER$^{-ve}$ disease (OR: 0.73; 95% CI: 0.43, 1.25) comparing the top vs bottom tertiles of SHBG concentrations.[6] A case-cohort analysis in the Melbourne Collaborative Cohort that included 132 ER$^{+ve}$ and 45 ER$^{-ve}$ women with breast cancer observed inverse associations for SHBG concentrations with both ER$^{+ve}$ [hazard ratio (HR) per doubling of SHBG: 0.41; 95% CI: 0.27, 0.63] and ER$^{-ve}$ disease (HR: 0.44; 95% CI: 0.23, 0.83).[4] Results from the Nurses' Health Study nested case-control study that included 147 women with ER$^+$/progesterone receptor positive (PR$^+$) postmenopausal breast cancer and 622 controls, yielded an OR of 0.50 (95% CI: 0.30, 0.80) comparing women in the highest versus the lowest quartile of SHBG concentrations. However, there was little evidence of associations for ER$^-$/PR$^-$ ($N = 38$ cases) and ER$^+$/PR$^-$ ($N = 33$ cases) disease.[8] No associations of circulating SHBG concentrations with ER$^{+ve}$ ($N = 127$ cases) and ER$^{-ve}$($N = 30$) postmenopausal breast cancer were recorded in the ORDET cohort.[9] Our MR investigation observed an inverse association between genetically determined SHBG concentrations

and risk of ER$^{+ve}$ breast cancer in agreement with the direction of the majority of the existing observational literature, but we also observed an increased risk for ER$^{-ve}$ disease, which is a novel finding and warrants further investigation given the wider observed variability in this analysis and the low prior probability of being true.[48]

Breast cancer is a complex and heterogeneous disease with a variety of histopathological and molecular subtypes that have diverse risk factors and clinical outcomes.[49] The associations of estrogens and androgens with a higher risk of postmenopausal ER$^{+ve}$ breast cancer are well established, but the literature is sparse and inconsistent for ER$^{-ve}$ disease. The observed positive association between genetically determined circulating SHBG concentrations and risk of ER$^{-ve}$ breast cancer, which was qualitatively different from the association observed for ER$^{+ve}$ disease, does not have a straightforward explanation, but it is biologically plausible given the pleiotropic actions of SHBG.[12,13] For many years, SHBG was believed to serve exclusively as a transporter or reservoir for sex steroids. However, in the past two decades it became clear that cell membranes of many tissues express a receptor for SHBG and that SHBG is found intracellularly.[12,13] Binding of SHBG to its receptor has been shown to activate cyclic adenosine monophosphate (cAMP),[13] an intracellular signal transduction pathway important for many biological processes including cancer growth.[50,51] It has been also shown that the ligand-bound SHBG receptor can activate the androgen receptor in the prostate in the absence of androgens.[52] Preclinical evidence indicates that testosterone has antiproliferative effects on mammary cell growth regulated by the androgen receptor.[53] A case-cohort study in the Women's Health Initiative Observational Study showed that higher serum concentrations of bioavailable testosterone were associated with lower risks of ER$^{-ve}$ postmenopausal breast cancer,[54] providing indirect evidence in accordance with our findings for SHBG and ER$^{-ve}$ disease. Additional studies on SHBG and ER$^{-ve}$ breast cancer are required to delineate potential mechanisms linking SHBG to this subtype.

This is the first study, to our knowledge, that investigated the potential causal association between SHBG concentrations with risk of overall breast cancer and cancer by ER status, overcoming the potential limitations of observational studies. Our MR study was powered to detect the effect sizes that we found. The F statistic was 208.9, assuming that the variance explained by the genetic instrument is approximately 8%, indicating a strong instrument. Nevertheless, several limitations should also be considered in interpreting our findings. MR estimates have a causal interpretation only if the assumptions of the method hold. Though it is not possible to prove the validity of some of these assumptions, we performed sensitivity analyses and used several statistical tests to

investigate potential violations. One out of the 12 variants associated with SHBG concentrations (i.e. rs12150660) has been also associated at a genome-wide significance level with testosterone concentrations in men,[55] but this variant is located in the *SHBG* gene, and will likely lead to vertical (not horizontal) pleiotropy, not violating thus the results of the present study.[56] In addition, most known genetic signals for estradiol and testosterone have only captured variability in men, precluding an MR analysis for these hormones in relation to breast cancer.[55,57–59] The summary-level data that we used did not allow for stratified analyses by covariates of interest, such as menopausal status, exogenous hormone use or according to breast cancer by progesterone and HER2 receptor status. Information on menopausal status was not available in the large genetic network that we used, but approximately 85% of breast cancer cases in the sample were postmenopausal at diagnosis.[24] Moreover, summary statistics for all genome-wide significant SNPs for BMI[60] were not available in the respective GWAS for SHBG,[20] and thus these could not be incorporated in a unified multivariable MR framework as was performed for other traits.[61] Consequently, the potential causal association of BMI with breast cancer cannot be quantified by this study or compared with estimates from another MR study.[62] Future large pooling consortia, genome-wide association studies of estradiol, testosterone and SHBG concentrations in women with expanded sample size, and MR studies with individual-level data could provide improved understanding of the role of sex steroids in breast tumorigenesis.

In summary, using a comprehensive MR analytical framework, we corroborated the previous literature evidence coming from observational studies for a potentially causal inverse association between SHBG concentrations and risk of ER$^{+ve}$ breast cancer. At the same time, our findings suggested a novel positive association with ER$^{-ve}$ disease, which warrants further investigation given the low prior probability of being true, but might indicate that the role of SHBG in breast cancer development is complex, exerting differential effects depending on ER status.

## Supplementary Data

Supplementary data are available at *IJE* online.

## Funding

## References

1. Collaborative Group on Hormonal Factors in Breast C. Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. *Lancet Oncol* 2012;**13**:1141–51.

2. Hankinson SE, Colditz GA, Willett WC. Towards an integrated model for breast cancer etiology: the lifelong interplay of genes, lifestyle, and hormones. *Breast Cancer Res* 2004;**6**:213–18.

3. Key T, Appleby P, Barnes I, Reeves G; Endogenous Hormones and Breast Cancer Collaborative Group. Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies. *J Natl Cancer Inst* 2002;**94**:606–16.

4. Baglietto L, Severi G, English DR *et al*. Circulating steroid hormone levels and risk of breast cancer for postmenopausal women. *Cancer Epidemiol Biomarkers Prev* 2010;**19**:492–502.

5. Cummings SR, Lee JS, Lui LY, Stone K, Ljung BM, Cauleys JA. Sex hormones, risk factors, and risk of estrogen receptor-positive breast cancer in older women: a long-term prospective study. *Cancer Epidemiol Biomarkers Prev* 2005;**14**:1047–51.

6. James RE, Lukanova A, Dossus L *et al*. Postmenopausal serum sex steroids and risk of hormone receptor-positive and -negative breast cancer: a nested case-control study. *Cancer Prev Res (Phila)* 2011;**4**:1626–35.

7. Kahan Z, Gardi J, Nyari T *et al*. Elevated levels of circulating insulin-like growth factor-I, IGF-binding globulin-3 and testosterone predict hormone-dependent breast cancer in postmenopausal women: a case-control study. *Int J Oncol* 2006;**29**:193–200.

8. Missmer SA, Eliassen AH, Barbieri RL, Hankinson SE. Endogenous estrogen, androgen, and progesterone concentrations and breast cancer risk among postmenopausal women. *J Natl Cancer Inst* 2004;**96**:1856–65.

9. Sieri S, Krogh V, Bolelli G *et al*. Sex hormone levels, breast cancer risk, and cancer receptor status in postmenopausal women: the ORDET cohort. *Cancer Epidemiol Biomarkers Prev* 2009;**18**:169–76.

10. Zeleniuch-Jacquotte A, Toniolo P, Levitz M *et al*. Endogenous estrogens and risk of breast cancer by estrogen receptor status: a

prospective study in postmenopausal women. *Cancer Epidemiol Biomarkers Prev* 1995;**4**:857–60.

11. Fortunati N, Catalano MG. Sex hormone-binding globulin (SHBG) and estradiol cross-talk in breast cancer cells. *Horm Metab Res* 2006;**38**:236–40.

12. Avvakumov GV, Cherkasov A, Muller YA, Hammond GL. Structural analyses of sex hormone-binding globulin reveal novel ligands and function. *Mol Cell Endocrinol* 2010;**316**:13–23.

13. Rosner W, Hryb DJ, Kahn SM, Nakhla AM, Romas NA. Interactions of sex hormone-binding globulin with target cells. *Mol Cell Endocrinol* 2010;**316**:79–85.

14. He XY, Liao YD, Yu S, Zhang Y, Wang R. Sex hormone binding globulin and risk of breast cancer in postmenopausal women: a meta-analysis of prospective studies. *Horm Metab Res* 2015;**47**:485–90.

15. Endogenous Hormones and Breast Cancer Collaborative Group. Sex hormones and risk of breast cancer in premenopausal women: a collaborative reanalysis of individual participant data from seven prospective studies. *Lancet Oncol* 2013;**14**:1009–19.

16. Kaaks R, Berrino F, Key T *et al*. Serum sex steroids in premenopausal women and breast cancer risk within the European Prospective Investigation into Cancer and Nutrition (EPIC). *J Natl Cancer Inst* 2005;**97**:755–65.

17. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003;**32**:1–22.

18. Ebrahim S, Davey Smith G. Mendelian randomization: can genetic epidemiology help redress the failures of observational epidemiology? *Hum Genet* 2008;**123**:15–33.

19. Prescott J, Thompson DJ, Kraft P *et al*. Genome-wide association study of circulating estradiol, testosterone, and sex hormone-binding globulin in postmenopausal women. *PLoS One* 2012;**7**:e37815.

20. Coviello AD, Haring R, Wellons M *et al*. A genome-wide association meta-analysis of circulating sex hormone-binding globulin reveals multiple loci implicated in sex steroid hormone regulation. *PLoS Genet* 2012;**8**:e1002805.

21. Burgess S, Scott RA, Timpson NJ, Davey Smith G, Thompson SG; EPIC-Interact Consortium. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. *Eur J Epidemiol* 2015;**30**:543–52.

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

23. Locke AE, Kahali B, Berndt SI *et al*. Genetic studies of body mass index yield new insights for obesity biology. *Nature* 2015;**518**:197–206.

24. Michailidou K, Lindstrom S, Dennis J *et al*. Association analysis identifies 65 new breast cancer risk loci. *Nature* 2017;**551**:92–94.

25. Rodriguez-Barranco M, Tobias A, Redondo D, Molina-Portillo E, Sanchez MJ. Standardizing effect size from linear regression models with log-transformed variables for meta-analysis. *BMC Med Res Methodol* 2017;**17**:44.

26. Brion MJ, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. *Int J Epidemiol* 2013;**42**:1497–501.

27. Rohrmann S, Shiels MS, Lopez DS *et al*. Body fatness and sex steroid hormone concentrations in US men: results from NHANES III. *Cancer Causes Control* 2011;**22**:1141–51.

28. Kyrgiou M, Kalliala I, Markozannes G *et al*. Adiposity and cancer at major anatomical sites: umbrella review of the literature. *BMJ* 2017;**356**:j477.

29. Lauby-Secretan B, Scoccianti C, Loomis D, Grosse Y, Bianchini F, Straif K. Body Fatness and Cancer - Viewpoint of the IARC Working Group. *N Engl J Med* 2016;**375**:794–98.

30. Rees JMB, Wood AM, Burgess S. Extending the MR-Egger method for multivariable Mendelian randomization to correct for both measured and unmeasured pleiotropy. *Stat Med* 2017;**36**:4705–18.

31. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* 2013;**37**:658–65.

32. Dastani Z, Hivert MF, Timpson N *et al*. Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: a multi-ethnic meta-analysis of 45, 891 individuals. *PLoS Genet* 2012;**8**:e1002607.

33. Greenland S. An introduction to instrumental variables for epidemiologists. *Int J Epidemiol* 2000;**29**:722–29.

34. Martens EP, Pestman WR, de Boer A, Belitser SV, Klungel OH. Instrumental variables: application and limitations. *Epidemiology* 2006;**17**:260–67.

35. Welter D, MacArthur J, Morales J *et al*. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucl Acids Res* 2014;**42**:D1001–06.

36. Bowden J, Del Greco MF, Minelli C, Davey Smith G, Sheehan N, Thompson J. A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Stat Med* 2017;**36**:1783–802.

37. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* 2015;**44**:512–25.

38. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol* 2016;**40**:304–14.

39. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol* 2017;**46**:1985–98.

40. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet* 2018;**50**:693–98.

41. Below JE, Parra EJ, Gamazon ER *et al*. Meta-analysis of lipid-traits in Hispanics identifies novel loci, population-specific effects, and tissue-specific enrichment of eQTLs. *Sci Rep* 2016;**6**:19429.

42. Eriksson N, Benton GM, Do CB *et al*. Genetic variants associated with breast size also influence breast cancer risk. *BMC Med Genet* 2012;**13**:53.

43. Franke A, McGovern DP, Barrett JC *et al*. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet* 2010;**42**:1118–25.

44. Yang Q, Kottgen A, Dehghan A *et al*. Multiple genetic loci influence serum urate levels and their relationship with gout and cardiovascular disease risk factors. *Circ Cardiovasc Genet* 2010;**3**:523–30.

45. Yavorska OO, Burgess S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int J Epidemiol* 2017;**46**:1734–39.

46. Spiller W, Davies N, Palmer T. Software Application Profile: mrrobust - A tool for performing two-sample summary Mendelian randomization analyses. bioRxiv 2017. doi:10.1101/142125.

47. Slob E, Burgess S, A comparison of robust mendelian randomization methods using summary data. bioRxiv 2019. doi:10.1101/577940.

48. Wakefield J. Reporting and interpretation in genome-wide association studies. *Int J Epidemiol* 2008;**37**:641–53.

49. Weigelt B, Reis-Filho JS. Histological and molecular types of breast cancer: is there a unifying taxonomy? *Nat Rev Clin Oncol* 2009;**6**:718–30.

50. Abramovitch R, Tavor E, Jacob-Hirsch J *et al.* A pivotal role of cyclic AMP-responsive element binding protein in tumor progression. *Cancer Res* 2004;**64**:1338–46.

51. Simpson BJ, Ramage AD, Hulme MJ *et al.* Cyclic adenosine 3', 5'-monophosphate-binding proteins in human ovarian cancer: correlations with clinicopathological features. *Clin Cancer Res* 1996;**2**:201–06.

52. Nakhla AM, Romas NA, Rosner W. Estradiol activates the prostate androgen receptor and prostate-specific antigen secretion through the intermediacy of sex hormone-binding globulin. *J Biol Chem* 1997;**272**:6838–41.

53. Dimitrakakis C, Bondy C. Androgens and the breast. *Breast Cancer Res* 2009;**11**:212.

54. Farhat GN, Cummings SR, Chlebowski RT *et al.* Sex hormone levels and risks of estrogen receptor-negative and estrogen receptor-positive breast cancers. *J Natl Cancer Inst* 2011;**103**:562–70.

55. Ohlsson C, Wallaschofski H, Lunetta KL *et al.* Genetic determinants of serum testosterone concentrations in men. *PLoS Genet* 2011;**7**:e1002313.

56. Hemani G, Bowden J, Davey Smith G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum Mol Genet* 2018;**27**:R195–R208.

57. Chen Z, Tao S, Gao Y *et al.* Genome-wide association study of sex hormones, gonadotropins and sex hormone-binding protein in Chinese men. *J Med Genet* 2013;**50**:794–801.

58. Eriksson AL, Perry JRB, Coviello AD *et al.* Genetic determinants of circulating estrogen levels and evidence of a causal effect of estradiol on bone density in men. *J Clin Endocrinol Metab* 2018;**103**:991–1004.

59. Jin G, Sun J, Kim ST *et al.* Genome-wide association study identifies a new locus JMJD1C at 10q21 that may influence serum androgen levels in men. *Hum Mol Genet* 2012;**21**:5222–28.

60. Yengo L, Sidorenko J, Kemper KE *et al.* Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. *Hum Mol Genet* 2018;**27**:3641–49.

61. Burgess S, Freitag DF, Khan H, Gorman DN, Thompson SG. Using multivariable Mendelian randomization to disentangle the causal effects of lipid fractions. *PLoS One* 2014;**9**:e108891.

62. Guo Y, Warren Andersen S, Shu XO *et al.* Genetically predicted body mass index and breast cancer risk: Mendelian randomization analyses of data from 145,000 women of European descent. *PLoS Med* 2016;**13**:e1002105.

# Commentary: What can Mendelian randomization tell us about causes of cancer?

## Daniela Mariosa,[1] Robert Carreras-Torres,[1,2] Richard M Martin,[3,4,5] Mattias Johansson[1] and Paul Brennan[1]*

[1]Section of Genetics, International Agency for Research on Cancer (IARC), Lyon, France, [2]ONCOBELL Program, Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de Llobregat, Spain, [3]MRC Integrative Epidemiology Unit, [4]Department of Population Health Sciences, Bristol Medical School, and [5]University Hospitals Bristol NHS Foundation Trust National Institute for Health Research Bristol, Nutrition Biomedical Research Unit, University of Bristol, Bristol, UK

*Corresponding author. Section of Genetics, International Agency for Research on Cancer, 150 Cours Albert Thomas, 69372 Lyon cedex 08, France. E-mail: brennanP@iarc.fr

Multiple important causes of cancer have been successfully identified over the past 70 years, including cigarette smoking, alcohol, obesity and UV light, as well as carcinogens in the occupational environment and different infections. However, despite these successes, about half of the cancer burden cannot be linked to known causes.[1] Difficulties in identifying causal factors for different cancers are due to a number of reasons, including limitations in epidemiological study designs and the inherent problems of confounding and reverse causation, as well as inadequate statistical power to study relatively rare cancer types. Potential causes may also be relatively ubiquitous within populations, such as air