



## Mendelian Randomization

# The genetic interplay between body mass index, breast size and breast cancer risk: a Mendelian randomization analysis

**Brandon Nick Sern Ooi<sup>1†</sup>, Huiwen Loh<sup>1†</sup>, Peh Joo Ho<sup>1</sup>, Roger L Milne<sup>2</sup>, Graham Giles<sup>2</sup>, Chi Gao<sup>3</sup>, Peter Kraft<sup>3</sup>, Esther M John<sup>4</sup>, Anthony Swerdlow<sup>5</sup>, Hermann Brenner<sup>6,7,8</sup>, Anna H Wu<sup>9</sup>, Christopher Haiman<sup>9</sup>, D Gareth Evans<sup>10</sup>, Wei Zheng <sup>11</sup>, Peter A Fasching<sup>12</sup>, Jose Esteban Castela<sup>13</sup>, Ava Kwong<sup>14</sup>, Xia Shen<sup>15,16,17</sup>, Kamila Czene<sup>13</sup>, Per Hall<sup>13</sup>, Alison Dunning<sup>18</sup>, Douglas Easton<sup>18</sup>, Mikael Hartman<sup>19</sup> and Jingmei Li<sup>1,19\*</sup>**

<sup>1</sup>Human Genetics, Genome Institute of Singapore, Singapore, Singapore, <sup>2</sup>Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, VIC, Australia, <sup>3</sup>Program in Genetic Epidemiology and Statistical Genetics, Harvard T.H. Chan School of Public Health Boston, USA, <sup>4</sup>Department of Medicine and Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA, <sup>5</sup>Division of Genetics and Epidemiology and Division of Breast Cancer Research, Institute of Cancer Research, London UK, <sup>6</sup>Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ) Heidelberg, Germany, <sup>7</sup>Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany, <sup>8</sup>German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany, <sup>9</sup>Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA, <sup>10</sup>Genomic Medicine, Division of Evolution & Genomic Sciences, The University of Manchester Manchester, UK, <sup>11</sup>Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt University Medical Centre, Vanderbilt University Nashville, USA, <sup>12</sup>Department of Gynecology and Obstetrics, University Hospital Erlangen, Comprehensive Cancer Center Erlangen-EMN, Friedrich-Alexander University Erlangen-Nuremberg, Germany, <sup>13</sup>Oncology and Genetics Unit, Instituto de Investigacion Sanitaria Galicia Sur (IISGS), Xerencia de Xestion Integrada de Vigo-SERGAS, Vigo, Spain, <sup>14</sup>Department of Surgery, The University of Hong Kong Pok Fu Lam, Hong Kong, <sup>15</sup>Department of Medical Epidemiology and Biostatistics, Karolinska Institute Stockholm, Sweden, <sup>16</sup>Biostatistics Group, State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-sen University, Guangzhou, China, <sup>17</sup>Center for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK, <sup>18</sup>Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK, and <sup>19</sup>Department of Surgery, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

\*Corresponding author. Human Genetics, Genome Institute of Singapore, 60 Biopolis Street, Genome, #02-01, Singapore 138672, Singapore. E-mail: [lijm1@gis.a-star.edu.sg](mailto:lijm1@gis.a-star.edu.sg)

<sup>†</sup>These authors contributed equally to this work.

Editorial decision 23 May 2019; Accepted 31 May 2019

## Abstract

**Background:** Evidence linking breast size to breast cancer risk has been inconsistent, and its interpretation is often hampered by confounding factors such as body mass index (BMI). Here, we used linkage disequilibrium score regression and two-sample Mendelian randomization (MR) to examine the genetic associations between BMI, breast size and breast cancer risk.

**Methods:** Summary-level genotype data from 23andMe, Inc (breast size,  $n = 33\,790$ ), the Breast Cancer Association Consortium (breast cancer risk,  $n = 228\,951$ ) and the Genetic Investigation of ANthropometric Traits (BMI,  $n = 183\,507$ ) were used for our analyses. In assessing causal relationships, four complementary MR techniques [inverse variance weighted (IVW), weighted median, weighted mode and MR-Egger regression] were used to test the robustness of the results.

**Results:** The genetic correlation ( $rg$ ) estimated between BMI and breast size was high ( $rg = 0.50$ ,  $P = 3.89 \times 10^{-43}$ ). All MR methods provided consistent evidence that higher genetically predicted BMI was associated with larger breast size [odds ratio ( $OR_{IVW}$ ): 2.06 (1.80–2.35),  $P = 1.38 \times 10^{-26}$ ] and lower overall breast cancer risk [ $OR_{IVW}$ : 0.81 (0.74–0.89),  $P = 9.44 \times 10^{-6}$ ]. No evidence of a relationship between genetically predicted breast size and breast cancer risk was found except when using the weighted median and weighted mode methods, and only with oestrogen receptor (ER)-negative risk. There was no evidence of reverse causality in any of the analyses conducted ( $P > 0.050$ ).

**Conclusion:** Our findings indicate a potential positive causal association between BMI and breast size and a potential negative causal association between BMI and breast cancer risk. We found no clear evidence for a direct relationship between breast size and breast cancer risk.

**Key words:** Breast size, breast cancer risk, body mass index, Mendelian randomization, LDSC regression, genetic epidemiology, genetic correlation

### Key Messages

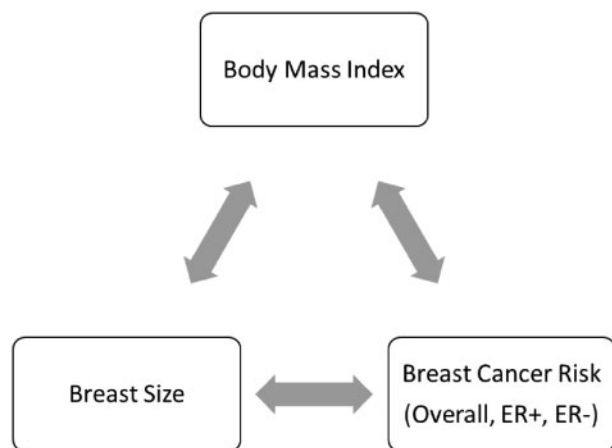
- The relationships between breast size, breast cancer risk and body mass index (BMI) were examined using summary-level genotype data from three publicly available sources.
- Genetic associations between these three factors were estimated using (i) genetic variants across the entire genome by linkage disequilibrium score regression, and (ii) only significantly associated variants as instrumental variables in two-sample Mendelian randomization analysis.
- Women with a genetic predisposition to high BMI are likely to have a larger breast size and a lower risk of breast cancer.
- There is no clear evidence for a direct and unmediated association between breast size and breast cancer risk.

## Introduction

Observational studies suggest that breast size is related to breast cancer risk, although the evidence is not consistent (reviewed in<sup>1</sup>). Some studies have shown that larger breast size may be associated with greater breast cancer risk<sup>2–4</sup> but others have reported an inverse relationship.<sup>5</sup> At the genetic level, Eriksson *et al.*<sup>6</sup> found that genetic variants associated with bigger bra cup size were also associated with increased breast cancer risk. However, the link between

breast size and breast cancer risk becomes less clear when other factors, such as body mass index (BMI), are considered at the same time.<sup>2,5</sup>

On the other hand, BMI has established relationships with both breast size and breast cancer risk. Generally, higher BMI is associated with larger breast size.<sup>7–9</sup> In a twin study, the overlap in genetic heritability for BMI and breast size was estimated to be ~33%.<sup>10</sup> However, the link between BMI and breast cancer risk is dependent on



**Figure 1.** Relationships between the three traits examined in this study.

menopausal status. An inverse relationship is commonly observed between BMI and premenopausal breast cancer; a positive relationship is commonly observed between BMI and postmenopausal breast cancer.<sup>11–14</sup> In terms of cancer subtype, the associations are reported to be stronger for hormone receptor-positive breast cancer in both pre- and post-menopausal women.<sup>14–16</sup> However, several Mendelian randomization (MR) analyses showed that BMI predicted by genome-wide association studies (GWAS)-identified variants was inversely associated with the risk of both pre- and post-menopausal breast cancer.<sup>17–19</sup>

In this study, we examined the genetic interplay between BMI, breast size and breast cancer risk (Figure 1). Specifically, we (i) used linkage disequilibrium score regression (LDSC)<sup>20</sup> to estimate genetic correlation by considering genetic variants across the entire genome, and (ii) assessed causality using only significantly associated genetic variants as instruments in two-sample MR analyses.<sup>21</sup> Both these methods have been selected for their respective strengths—LDSC for the case of complex traits where thousands of variants can have small effects, and MR for its ability to infer causality given that certain assumptions hold.

## Methods

### Summary statistics for BMI, breast size and breast cancer risk

Beta coefficients and standard errors for BMI, breast size and breast cancer risk for women of European ancestry were obtained from three publicly available sources: Genetic Investigation of ANthropometric Traits (GIANT),<sup>22</sup> 23andMe (version 4.1)<sup>23</sup> and the Breast Cancer Association Consortium (BCAC).<sup>24–26</sup> The genetic associations with BMI are in standard deviation (SD) units whereas breast size was coded from 0 to 9, corresponding

to an increase in units of bra cup size: smaller than AAA, AAA, AA, A, B, C, D, DD, DDD, and larger than DDD. Breast cancer was coded as a binary phenotype for overall, oestrogen receptor (ER)-positive and ER-negative breast cancer. Further details are provided in [Supplementary Table 1](#), available as [Supplementary data](#) at *IJE* online.

### Cross-trait genetic correlation estimated using LDSC

LDSC, which relies on the adjustment for linkage between single nucleotide polymorphisms (SNPs), was used to estimate cross-trait genetic correlations between BMI, breast size and breast cancer risk (overall, ER-positive, and ER-negative).<sup>20</sup> LDSC estimates genetic correlation by considering the effects of all SNPs, including those that do not reach genome-wide significance. SNPs for BMI, breast size and breast cancer risk, respectively, were merged with HapMap3 SNPs and duplicate reference SNP cluster ids (rsids) were removed using the software provided at <https://github.com/bulik/ldsc> ([Supplementary Table 1](#), available as [Supplementary data](#) at *IJE* online). Linkage disequilibrium scores provided by the software's creators, based on the 1000 Genome Project's European samples, were downloaded from the same website and cross-trait genetic correlation,  $rg$ , was computed from common SNPs between each pair of traits. After all preprocessing steps, there were 1 053 312 common SNPs for BMI and breast size, 992 833 common SNPs for BMI and all three breast cancer traits, and 1 116 435 common SNPs for breast size and all three breast cancer traits.

### SNP selection

Independent SNPs with genome-wide significant associations for BMI ( $n=77$ , explaining 2.2% of the variance in BMI)<sup>27</sup> and breast size ( $n=7$ , explaining ~1.2% of the variance in breast size)<sup>6</sup> were used as exposure data for MR analyses. For outcome, we included 172 breast cancer SNPs.<sup>24</sup> After verifying the independence of selected SNPs using the clumping function in the TwoSampleMR package in R,<sup>21</sup> 77, 7 and 114 SNPs were retained for further analyses, respectively ([Supplementary Tables 2–4](#), available as [Supplementary data](#) at *IJE* online). All 77 BMI SNPs were present in the breast size and breast cancer risk datasets. All 7 breast size SNPs were present in the breast cancer risk dataset, but only 4 were present in the BMI dataset. A total of 113 breast cancer SNPs were present in the breast size dataset and 81 SNPs were present in the BMI dataset. Before performing MR analyses, the signs of the beta coefficients for each SNP across all outcome datasets were aligned to the effective alleles of the exposure datasets.

To assess the MR assumptions, the MR-Egger test was used to detect directional pleiotropy of the genetic

instruments, where a regression intercept that significantly differs from zero indicates the presence of directional pleiotropy or the InSIDE assumption (INstrument Strength Independent of Direct Effect) was violated (i.e. the associations with other traits that may affect the outcome via pathways independent of the exposure).<sup>28</sup> For the datasets used in this study, only for the case of BMI (exposure) versus breast cancer risk (outcome) was the *P*-value for the MR-Egger intercept significantly different from zero. The MR-Pleiotropy RESidual Sum and Outlier (MR-PRESSO) package in R was used for pleiotropy and outlier SNPs detection for this dataset.<sup>29</sup> Six outliers were identified for BMI (exposure) and overall risk (outcome), four outliers for BMI (exposure) and ER-positive risk (outcome) and three outliers for BMI and ER-negative risk (outcome). After removal of these outliers, the MR-Egger regression intercept no longer differed significantly from zero. The MR-PRESSO package was also used to test the other datasets, but the MR results did not significantly differ when outliers identified in these other datasets were removed.

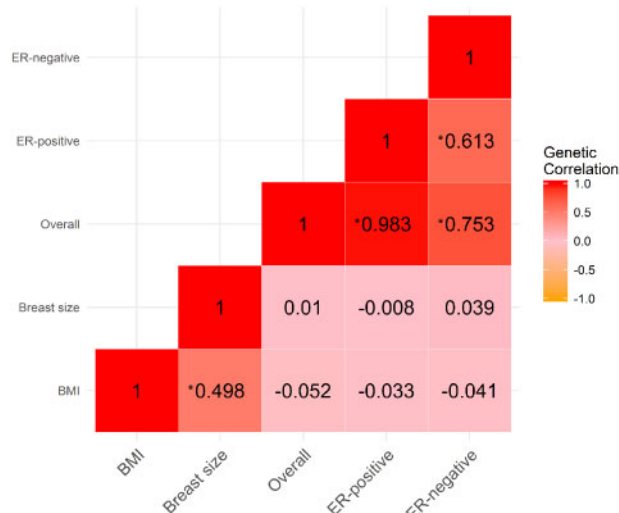
### Mendelian randomization

MR was first performed using the inverse-variance weighted (IVW) method.<sup>30</sup> The IVW method assumes that the MR assumptions are satisfied or that all SNPs are valid instruments. As it is difficult to test the validity of this assumption, three additional MR methods were also used to assess the robustness of the result from the primary analysis under alternative assumptions: weighted median function,<sup>31</sup> weighted mode function<sup>32</sup> and MR-Egger regression.<sup>28</sup> According to Hartwig *et al.*,<sup>32</sup> the weighted median function provides a valid result under the assumption that >50% of the weight in the model comes from SNPs that satisfy the MR assumptions. The weighted mode function is valid if the majority of SNPs with similar individual causal effect estimates are valid instruments even if other SNPs in the model do not meet the requirements for causal inference using MR. MR-Egger regression allows for horizontal pleiotropic effects. As long as this pleiotropy is not correlated with SNP-exposure associations, the beta-coefficient from the MR-Egger analysis is valid even if horizontal pleiotropy exists. Both forward and reverse causality for each pair of variables were investigated. The above statistical analyses were performed using TwoSampleMR package in R.<sup>21</sup>

## Results

### BMI is genetically correlated with breast size

Genetically predicted BMI and genetically predicted breast size were correlated [correlation coefficient (*rg*) = 0.50,

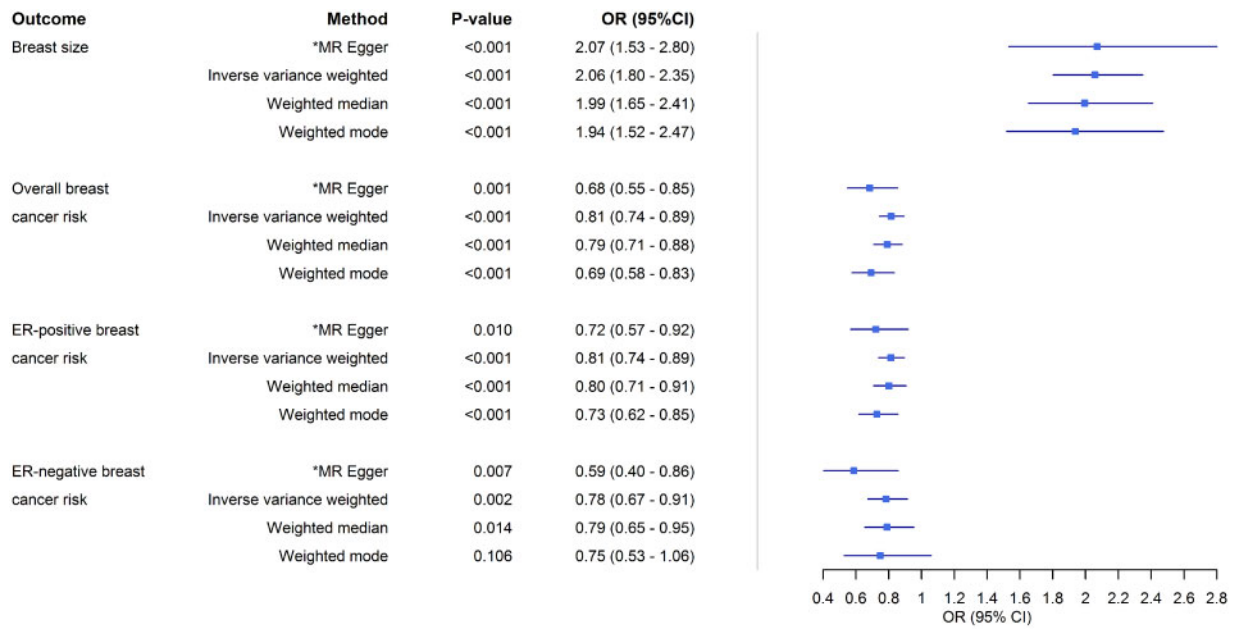


**Figure 2.** Genetic correlations between body mass index (BMI), breast size and breast cancer risk (overall, ER-positive and ER-negative). Colour intensity indicates correlation strength, with red indicating positive correlation (from 0 to +1) and yellow indicating negative correlation (from 0 to -1). Significant relationships ( $P < 0.05$ ) are denoted by asterisks.

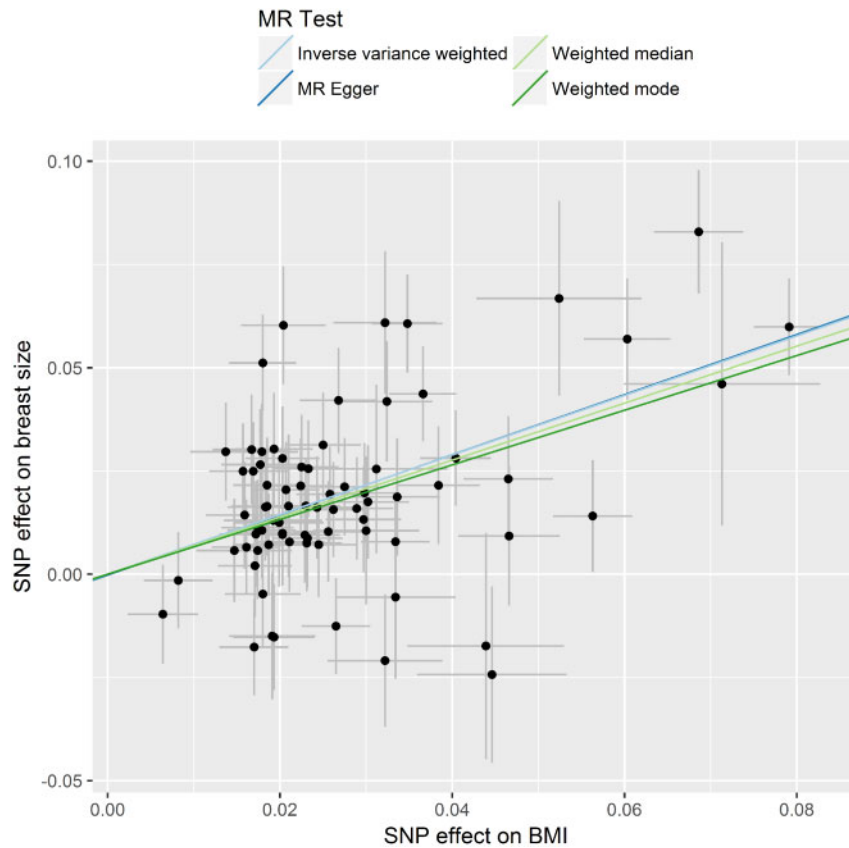
$P = 3.89 \times 10^{-43}$ ] (Figure 2 and Supplementary Table 5, available as Supplementary data at *IJE* online) but none of the other pairs of genetically predicted variables (BMI and breast cancer risk, breast size and breast cancer risk) was found by LDSC to be correlated ( $P > 0.050$ ). As expected, the genetic correlations between the three types of breast cancer (overall, ER-positive, ER-negative) were high ( $rg > 0.60$ , all  $P < 10^{-85}$ ) (Figure 2 and Supplementary Table 5, available as Supplementary data at *IJE* online).

### BMI has a positive effect on breast size but not vice versa

A 1-SD increase in genetically predicted BMI was strongly associated with an increase in genetically predicted breast size {odds ratio ( $OR_{IVW}$ ) [95% confidence interval (CI)]: 2.06 (1.80 – 2.35),  $P = 1.38 \times 10^{-26}$ , Figure 3}. This result was corroborated by other MR methods, namely, weighted median [ $OR_{\text{weighted median}}$ : 1.99 (1.64 – 2.42),  $P = 3.29 \times 10^{-12}$ ], weighted mode [ $OR_{\text{weighted mode}}$ : 1.94 (1.49–2.51),  $P = 3.82 \times 10^{-6}$ ] and MR-Egger [ $OR_{\text{MR-Egger}}$ : 2.07 (1.53 – 2.80),  $P = 1.04 \times 10^{-5}$ ] (Figures 3 and 4). The MR-Egger intercept test ( $P = 0.964$ ) suggested an absence of strong directional pleiotropy and leave-one-out permutation analysis (Figure 5) did not detect any single SNP that had a strong influence on the results. For the reverse direction, regression coefficients close to zero with non-significant *P*-values were obtained when breast size was used as exposure (Supplementary Figures 1 and 2, available as Supplementary data at *IJE* online).



**Figure 3.** Odds ratios (ORs) and 95% confidence intervals (CI) for the association between the exposure body mass index (BMI) and two outcomes (breast size and breast cancer risk) based on the different Mendelian randomization approaches used in this study. \*Value based on causal effect estimate from MR-Egger regression; corresponding MR-Egger intercept value testing presence of directional (bias inducing) pleiotropy not shown.



**Figure 4.** Scatter plot of SNP-breast size associations against SNP-body mass index (BMI) associations with estimates from different Mendelian randomization methods indicated by corresponding coloured lines.



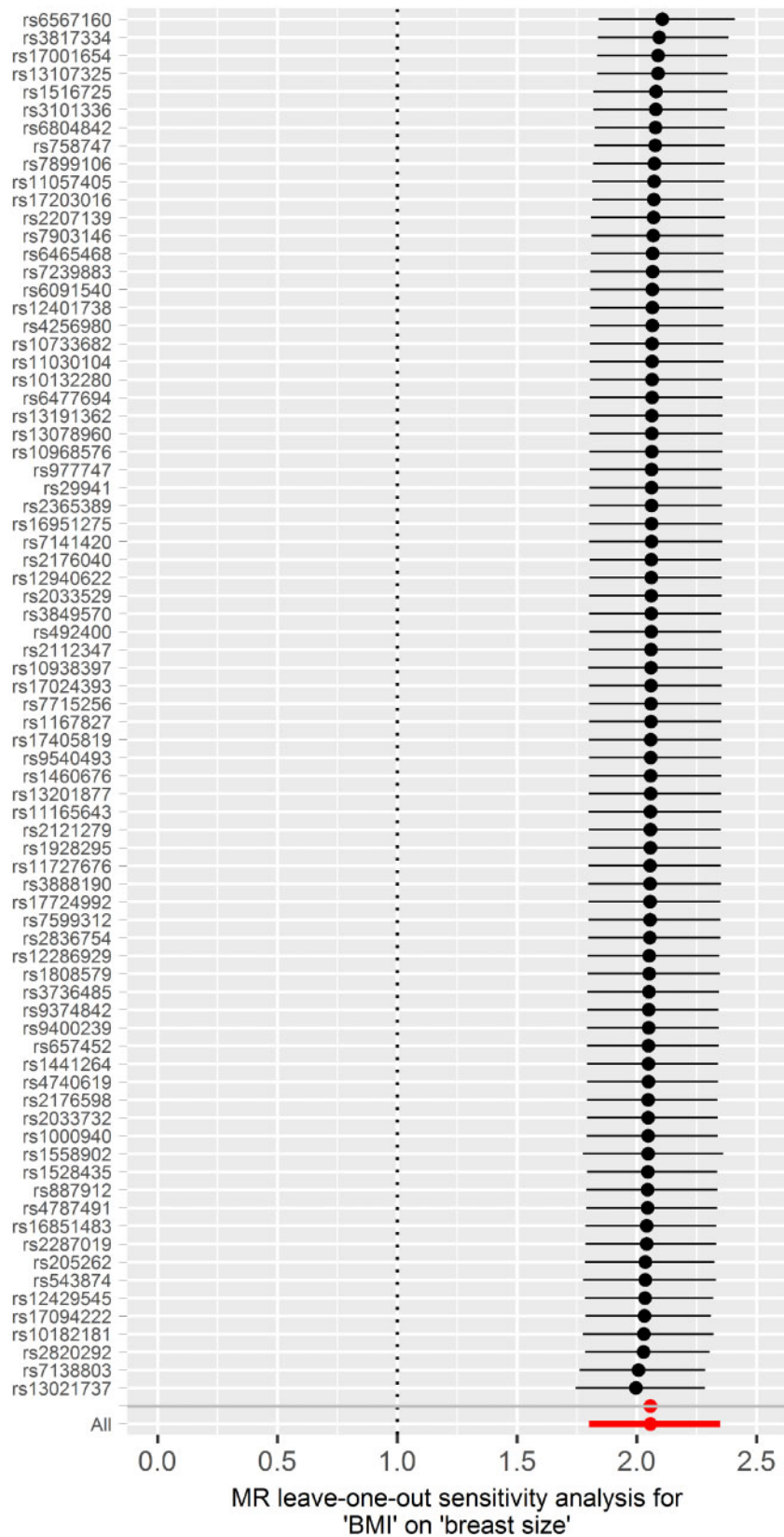
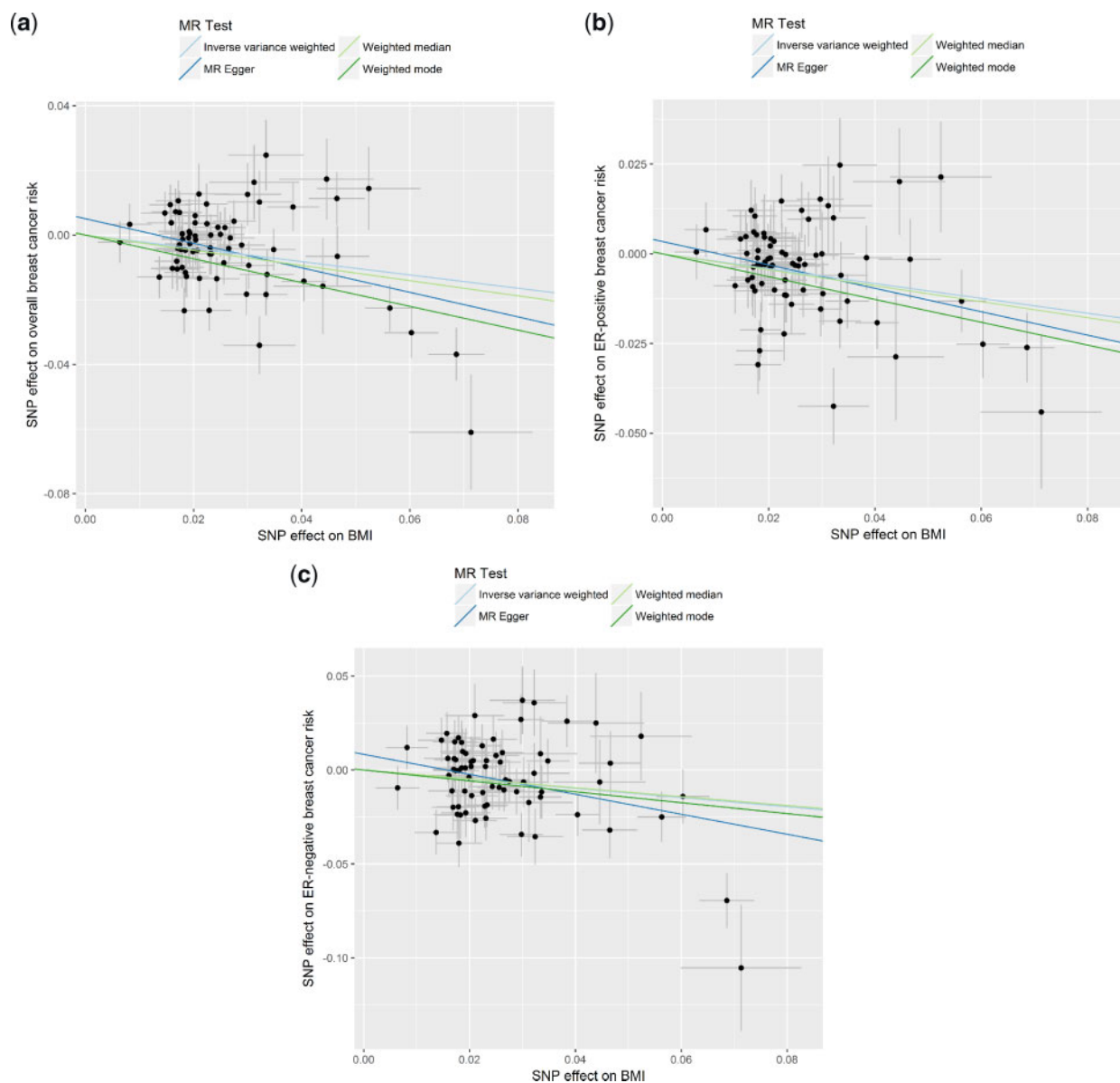


Figure 5. Leave-one-out permutation analysis plot for breast size obtained by leaving out the SNP indicated and repeating the standard inverse-variance weighted method with the rest of the 76 SNP instrumental variables.

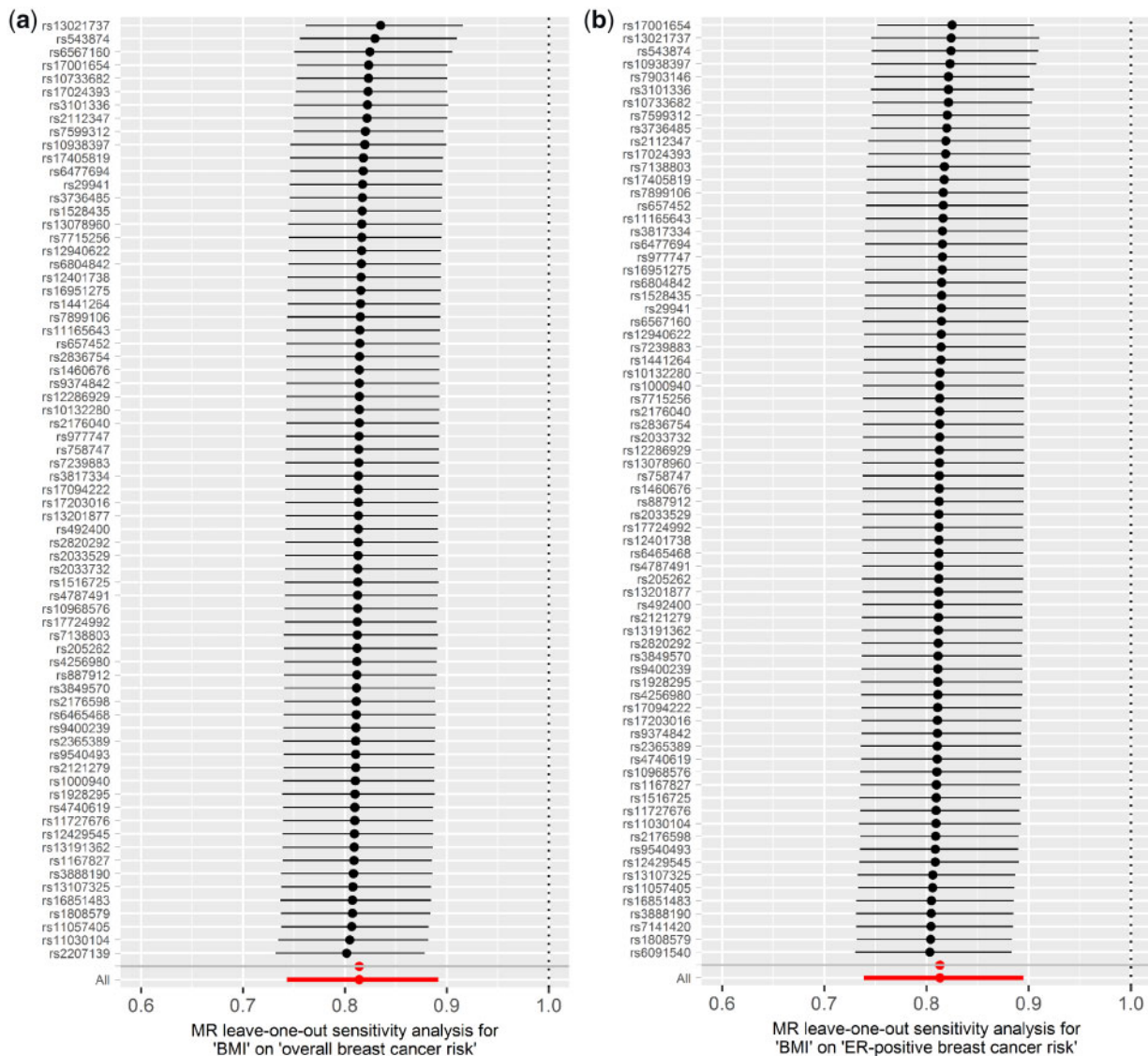


**Figure 6.** Scatter plots of SNP-breast cancer risk associations against SNP-body mass index (BMI) associations for (a) overall breast cancer risk, (b) oestrogen receptor (ER)-positive breast cancer risk, and (c) ER-negative breast cancer risk with estimates from different Mendelian randomization methods indicated by corresponding coloured lines.

### BMI has a negative effect on breast cancer risk but not vice versa

In the IVW analysis, an increase in genetically predicted BMI was strongly associated with a decrease in overall [OR<sub>IVW</sub>: 0.81 (0.74–0.89),  $P = 9.44 \times 10^{-6}$ ], ER-positive [OR<sub>IVW</sub>: 0.79 (0.72–0.88),  $P = 4.84 \times 10^{-6}$ ] and ER-negative [OR<sub>IVW</sub>: 0.78 (0.67–0.91),  $P = 1.78 \times 10^{-3}$ ] breast cancer risk (Figure 3). Similarly, the estimates derived from other MR methods did not appreciably differ from the IVW method (Figures 3 and 6a–c). No evidence for directional horizontal pleiotropy was observed ( $P = 0.095$ , 0.219 and 0.108 for BMI-overall, BMI-ER-positive and BMI-ER-

negative breast cancer risk, respectively) and none of the SNPs had an undue effect on the results (Figure 7a–c). These results were obtained using 71 SNPs (BMI-overall risk), 73 SNPs (BMI-ER-positive risk) and 74 SNPs (BMI-ER-negative risk) after removal of outliers as identified by MR-PRESSO. Before removal of outliers, the MR-Egger intercepts differed significantly from zero ( $P = 0.006$ , 0.026 and 0.002, respectively). For the reverse direction, regression coefficients close to zero were obtained when all three risks were used as exposure, and all  $P$ -values were not significant (all  $P > 0.070$ ) (Supplementary Figures 3b, 4b, 5b and 6A–C, available as [Supplementary data](#) at *IJE* online). One of



**Figure 7.** Leave-one-out permutation analysis plots for (a) overall breast cancer risk, (b) oestrogen receptor (ER)-positive breast cancer risk, and (c) ER-negative breast cancer risk obtained by leaving out the SNP indicated and repeating the standard inverse-variance weighted method with the rest of the 70, 72 and 73 SNP instrumental variables used respectively.

the SNPs (rs17817449) appeared to have a stronger effect compared with the other SNPs from the leave-one-out analysis (Supplementary Figure 6D–F, available as Supplementary data at *IJE* online), and this same outlier was also detected through MR-PRESSO, but removing it had a negligible effect on the results.

#### Unclear relationship between breast size and breast cancer risk

An increase in genetically predicted breast size was not significantly associated with breast cancer risk [OR<sub>IW</sub>: 1.23 (0.78–1.95),  $P=0.370$ , 1.19 (0.81–1.74),  $P=0.373$  and 1.32 (0.69–2.55),  $P=0.405$  for overall, ER-positive and

ER-negative breast cancer risk, respectively] (Figure 8). The OR estimates obtained from other methods were also not significant except for those obtained from the weighted median and weighted mode methods for the breast size-ER-negative association [OR<sub>weighted median</sub>: 1.32 (1.07–1.61),  $P=0.008$  and OR<sub>weighted mode</sub>: 1.44 (1.19–1.75),  $P=0.009$ ] (Figures 8 and 9a–c). No directional pleiotropic effect was detected as MR-Egger intercept tests did not differ significantly from zero ( $P=0.991$ , 0.891 and 0.836 for size-overall, size-ER-positive and size-ER-negative breast cancer risk, respectively), and no outliers were detected by leave-one-out permutation analysis (Figure 10a–c). For the reverse direction, regression coefficients close to zero were obtained when all three breast



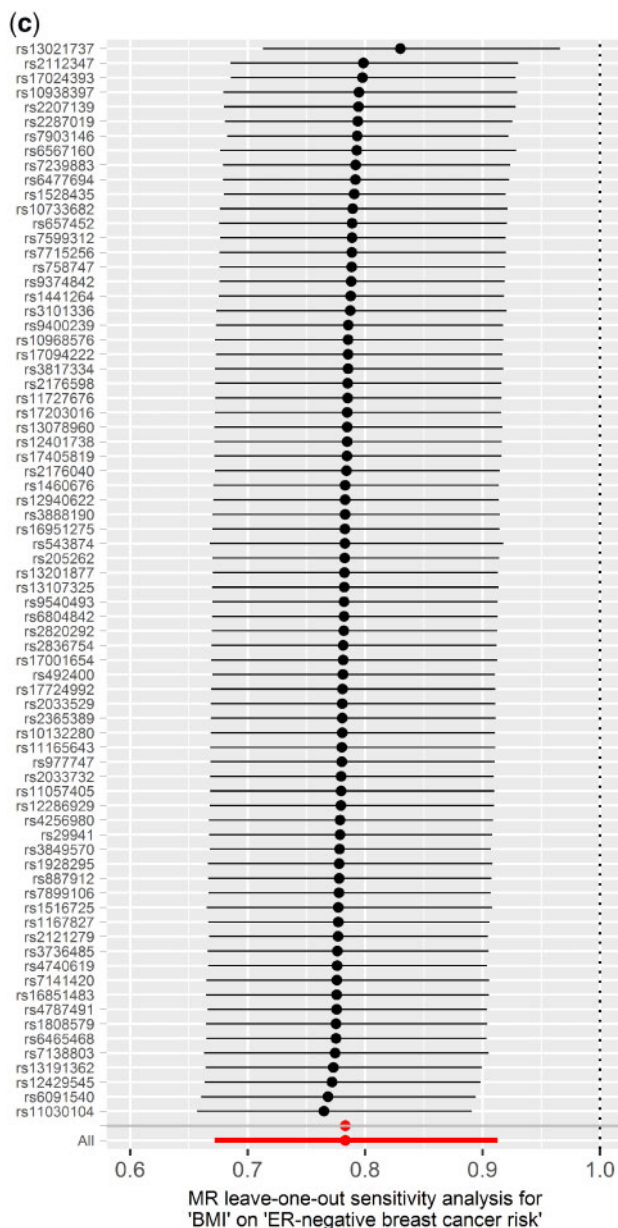


Figure 7. Continued.

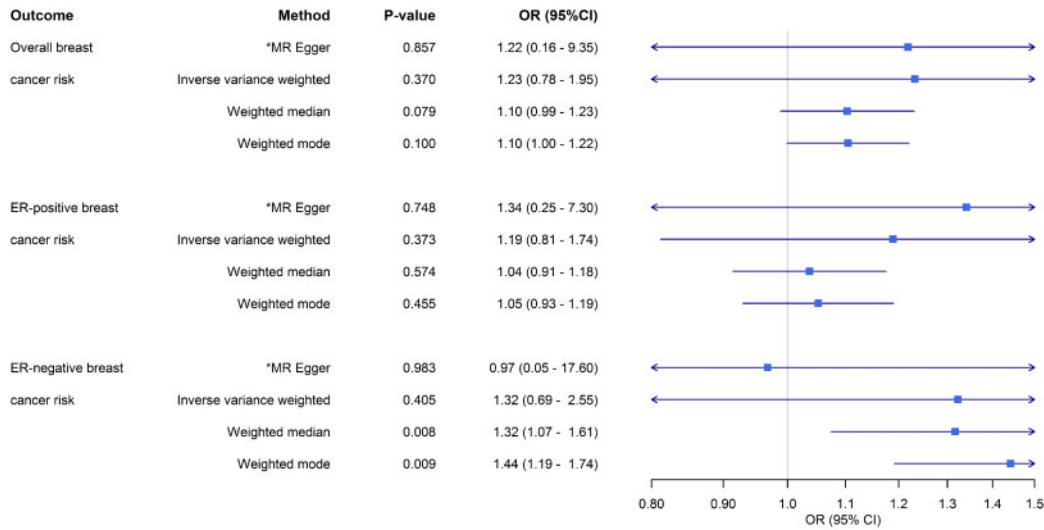
cancer risks were used as exposure and all  $P$ -values were not significant (all  $P > 0.470$ ) (Supplementary Figures 3a, 4a, 5a and 7, available as Supplementary data at *IJE* online).

One SNP (rs7816345) was reported to be associated with both BMI and breast density, which may undermine the validity of the instrument for breast size.<sup>6,33</sup> In a sensitivity analysis, removing SNP rs7816345 from the MR analysis resulted in attenuated results. In other words, the effect of breast size on ER-negative breast cancer risk was no longer significant [OR<sub>weighted median</sub>: 1.16 (0.95–1.43), and OR<sub>weighted mode</sub>: 0.94 (0.73–1.23)].

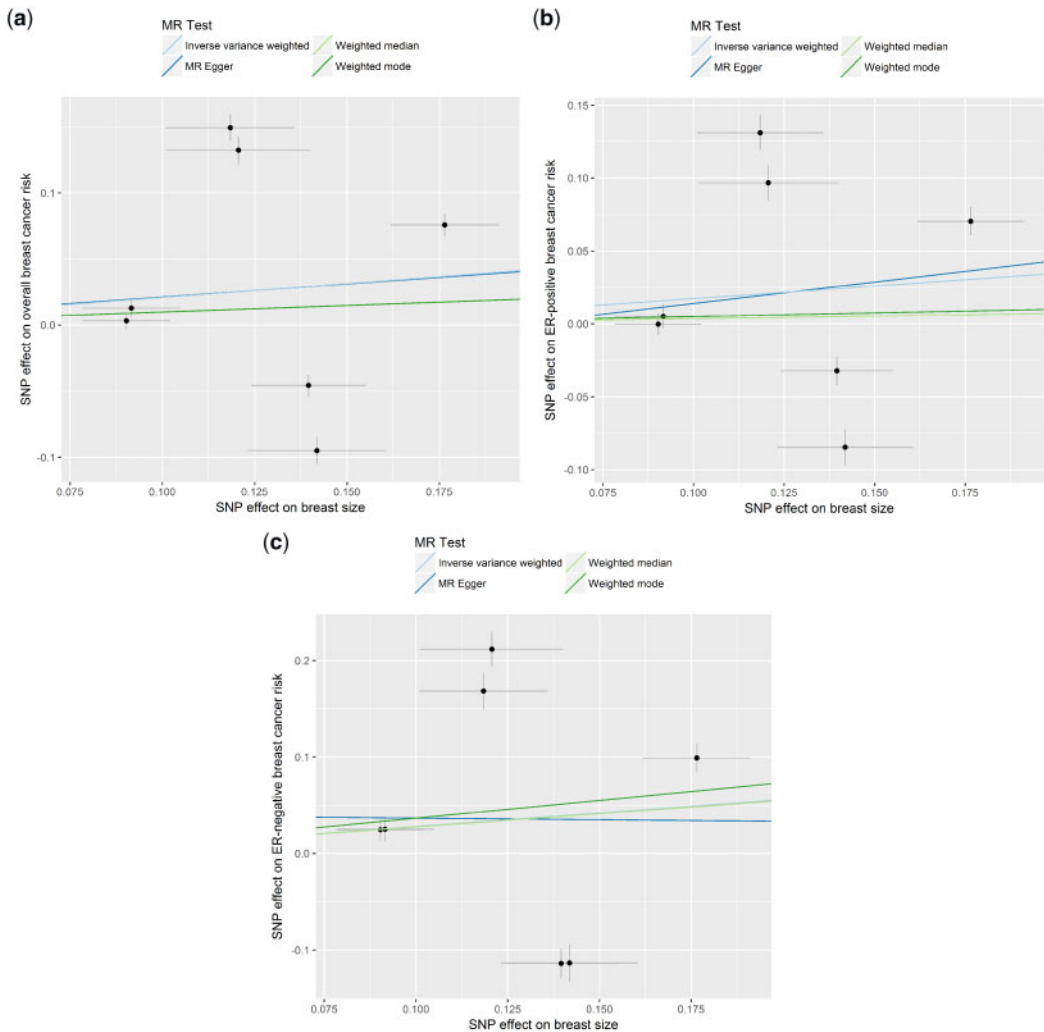
## Discussion

In this study, we found potential causal genetic evidence linking BMI to both breast size and breast cancer risk (Figure 11). Our finding that genetically predicted BMI is positively correlated with genetically predicted breast size is not surprising given that a number of other epidemiological studies have reported similar associations. Our analyses extend the finding of Wade *et al.*,<sup>10</sup> who used data from twin studies to infer that one third of genes contributing to breast size were in common with genes influencing BMI. Here, we used directly all SNPs in common between the two traits to calculate their genetic correlation after adjusting for genetic linkage, and found this correlation to be large and statistically significant. To our knowledge, this is the first time that empirical genetic data have been used directly to show a positive association between BMI and breast size. Additionally, our MR results suggest that a genetic predisposition to higher BMI may be causally linked to larger breast size, but not the reverse. This direction of association implies that women with higher BMI have larger breast size.

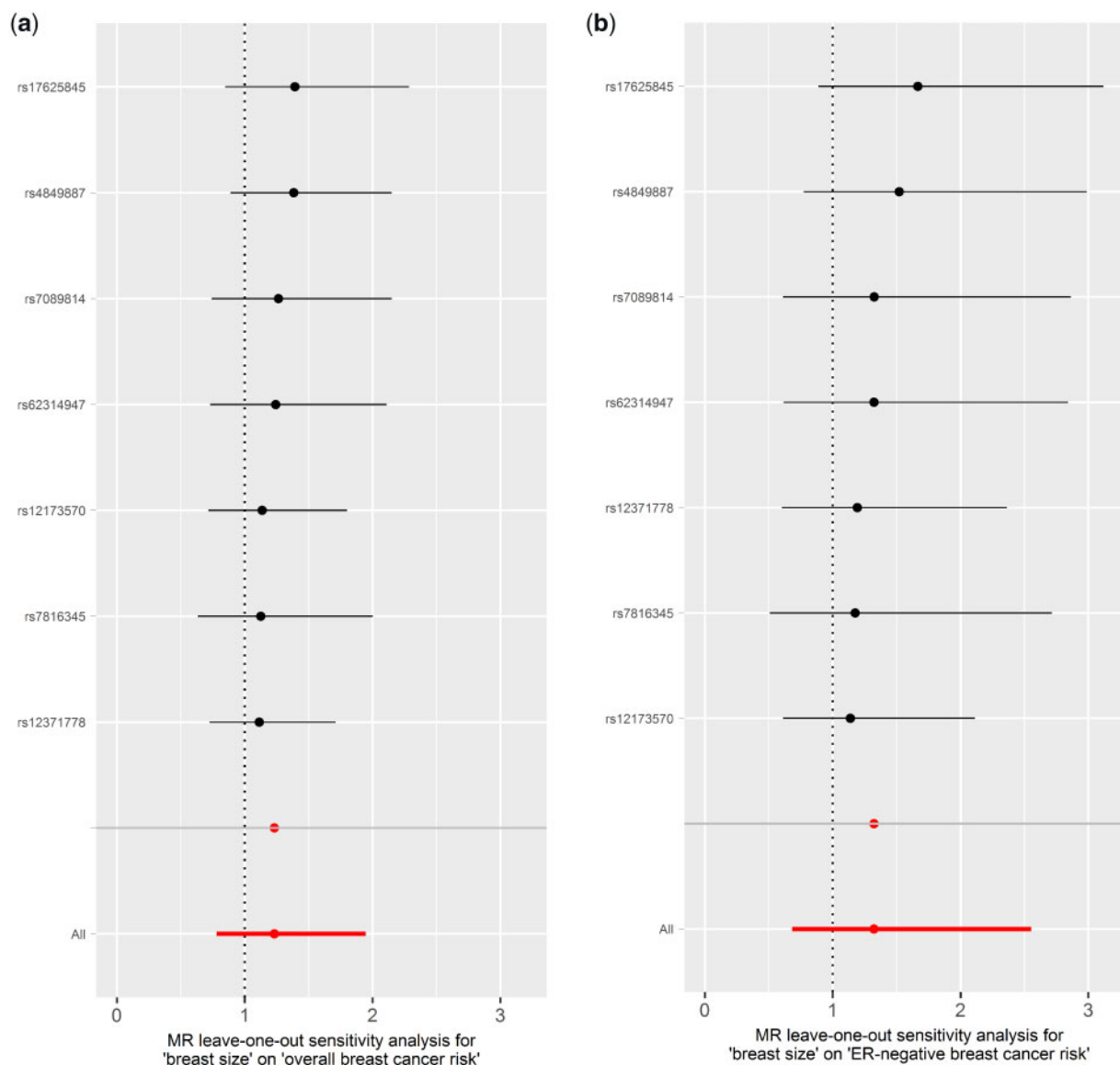
We also found through MR analyses a significant relationship between genetically predicted BMI and breast cancer risk. Higher genetically predicted BMI was found to be inversely associated with the risk of developing both ER-positive and ER-negative breast cancers. Some studies have suggested that BMI is more strongly associated with hormone receptor-positive cancers,<sup>14,15</sup> but we did not observe this. Furthermore, although ER-negative breast cancers are more common in premenopausal women, we did not find any differences in the (negative) association between BMI and breast cancer by ER-status. This contradicts the findings from other observational studies that suggest that BMI is positively associated with postmenopausal breast cancer but negatively associated with premenopausal breast cancer.<sup>11–13</sup> It is important to note that the GWAS summary statistics we used did not directly differentiate between pre- and post-menopausal breast cancer cases. Nevertheless, our findings are consistent with the MR analyses of Shu *et al.*<sup>19</sup> Gao *et al.*<sup>18</sup> and Guo *et al.*<sup>17</sup> who, using genetic data, found that BMI was inversely associated with both pre- and post-menopausal breast cancer risk. Given that other studies using a similar methodology have reported this inverse association and that it was independent of age, we were not overly surprised by the results. The protective effect obtained in our study (OR = 0.81) was weaker than that found by Guo *et al.* (OR = 0.65) but this difference could be due to the different datasets used as well as Guo *et al.*'s use of a genetic risk score as an instrumental variable instead of the beta-coefficients of each SNP that we used in our study.



**Figure 8.** Odds ratios (ORs) and 95% confidence intervals (CI) for the association between breast size and overall breast cancer risk, ER-positive breast cancer risk, and ER-negative breast cancer risk based on the different Mendelian randomization approaches used in this study. \*Value based on causal effect estimate from MR-Egger regression; corresponding MR-Egger intercept value testing presence of directional (bias inducing) pleiotropy not shown.



**Figure 9.** Scatter plots of SNP-breast cancer risk associations against SNP-breast size associations for (a) overall breast cancer risk, (b) oestrogen receptor (ER)-positive breast cancer risk, and (c) ER-negative breast cancer risk with estimates from different Mendelian randomization methods indicated by corresponding coloured lines.



**Figure 10.** Leave-one-out permutation analysis plots for (a) overall breast cancer risk, (b) oestrogen receptor (ER)-positive breast cancer risk, and (c) ER-negative breast cancer risk obtained by leaving out the SNP indicated and repeating the standard inverse-variance weighted method with the six SNP instrumental variables used respectively.

The positive association between BMI and breast cancer risk in postmenopausal women has been speculated to result from the higher tissue concentrations of oestrogen derived from the larger fat reserves of women with higher BMI.<sup>34</sup> On the other hand, the negative relationship between BMI and breast cancer risk in premenopausal women might be explained by lower levels of progesterone and oestrogen due to the longer anovulatory cycles that they experience.<sup>35,36</sup> In explaining why genetically predicted BMI was inversely correlated with postmenopausal risk, Guo *et al.*<sup>17</sup> suggest that the genetic portion of BMI may reflect an early-life BMI, or that weight gain during later adulthood and not BMI per se, could be the main factor leading to increased postmenopausal risk. These

reasons could also explain the results obtained from our MR analyses. Interestingly, whereas a negative correlation was estimated between BMI and breast cancer risk using LDSC, it was not significant (Supplementary Table 5, available as Supplementary data at *IJE* online). Also, MR analysis using breast cancer risk as the exposure and BMI as the outcome did not produce any significant associations. Thus, the association between BMI and breast cancer risk is mediated by only a limited number of SNPs that are significant for BMI but not for breast cancer risk.

Breast size is not a commonly considered breast cancer risk factor.<sup>1</sup> Indeed, the relationship between our final pair of traits, breast size and breast cancer risk, is less clear.

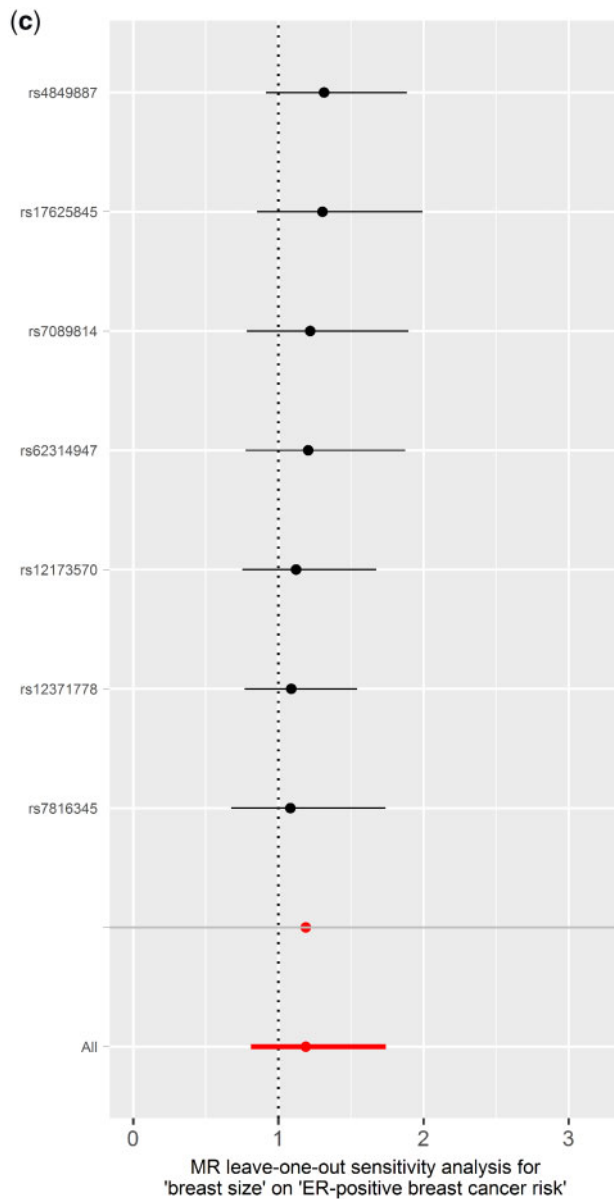


Figure 10. Continued.

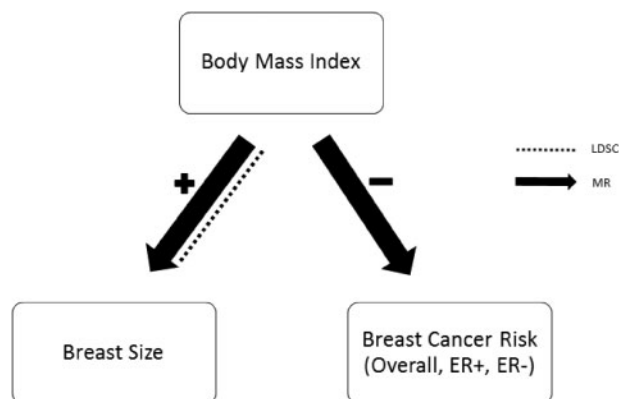


Figure 11. Relationships discovered in this study. Significant results from LDSC regression are denoted by dotted lines and significant results from MR analysis are denoted by solid arrows.

Eriksson *et al.*<sup>6</sup> reported two SNPs associated with breast size that were also associated with breast cancer risk, but when we used all seven SNPs for MR analysis, we were unable to detect a significant correlation except when using the weighted median and weighted mode methods. For these two methods, there was a positive relationship only between breast size and ER-negative breast cancer risk. The weighted median and weighted mode methods work on the assumption that some of the instrumental variables do not satisfy the requirements for valid MR analysis, but given that only seven SNPs were used for the analysis, this result has to be interpreted with caution. In addition, the average pleiotropy was not significantly different from 0, which may imply that there is no directional pleiotropy and that the estimate from the IVW method is not biased (i.e. results were consistent with the weighted median and weighted mode methods). Furthermore, when we used all SNPs for LDSC regression, no relationship was found as well. Taken together, there is no clear genetic evidence for a relationship between breast size and breast cancer risk.

Indeed, the relationship between breast size and breast cancer risk may be difficult to tease apart without the consideration of mammographic density, which is a strong, independent risk factor of breast cancer.<sup>37</sup> Total breast size (i.e. area) on a mammogram is a combination of radiologically dense and non-dense components. Absolute non-dense area on a mammogram is highly correlated with total breast area and is inversely associated with breast cancer risk.<sup>38</sup> Radiologically dense tissue, on the other hand, contributes to breast cancer risk.<sup>38</sup> It is unclear whether the breast size instrument currently used is catching size, density or a mixture of both. Future MR analyses examining relationships between mammographic density measures (i.e. percentage mammographic density, absolute dense area and absolute non-dense area) and breast cancer risk will be meaningful.

It is important to note other limitations to our analysis. Firstly, due to the use of summary statistics, we were not able to stratify our analyses based on BMI or other breast cancer risk factors. This could have provided further insight into the nature of the relationship between these two traits. Secondly, the association between breast size and breast cancer risk could be driven by environmental factors that have no genetic component making it undetectable in this study. Thirdly, the seven SNPs were identified using a semi-continuous measurement of bra cup size to measure breast size.<sup>6</sup> The lower variability in bra cup size as compared with a continuous variable, such as breast area, may have resulted in fewer SNPs being identified. However, the amount of variation in a phenotypic trait that is explained by genetic variants used in MR analyses is typically small (<1%, compared with ~1.2% for breast size).<sup>39</sup>

Our study is the first to address the potentially causal relationship between breast size and breast cancer using a MR approach. In summary, our results suggest that women with a genetic predisposition to high BMI are likely to have a larger breast size and a lower risk of breast cancer, but there is no clear evidence for a direct and unmediated association between breast size and breast cancer risk. Interestingly, the reverse causation of breast cancer risk to BMI, breast cancer risk to breast size and breast size to BMI were all not significant, which is in agreement with our intuition. The hypothesis that breast cancer risk is lower in women with both a larger breast size as well as a genetic predisposition for high BMI can be tested in future stratified studies.

## Supplementary Data

Supplementary data are available at *IJE* online.

## Funding

This work was supported by the National Research Foundation Singapore Fellowship (NRF-NRFF2017-02) awarded to J.L. D.G.E is a NIHR Senior investigator. The body mass index association data were obtained from the Genetic Investigation of ANthropometric Traits (GIANT) consortium, whereas the breast cancer genome-wide association meta-analyses, which generated the summary statistics used here, were supported by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research, the 'Ministère de l'Économie, de la Science et de l'Innovation du Québec' through Genome Québec and grant PSR-SIIRI-701, The US National Institutes of Health (U19 CA148065, X01HG007492), Cancer Research UK (C1287/A10118, C1287/A16563, C1287/A10710) and The European Union (HEALTH-F2-2009-223175 and H2020 633784 and 634935). All studies and funders for the breast cancer genome-wide association analyses are listed in Michailidou *et al.*<sup>24</sup>

## Acknowledgements

The data on breast size was obtained from 23andMe, Inc. and we would like to thank all research participants and employees of 23andMe for making this work possible.

**Conflict of interest:** None declared.

## References

- Jansen LA, Backstein RM, Brown MH. Breast size and breast cancer: a systematic review. *J Plast Reconstr Aesthet Surg* 2014; **67**:1615–23.
- Hsieh CC, Trichopoulos D. Breast size, handedness and breast cancer risk. *Eur J Cancer* 1991; **27**:131–35.
- Kusano AS, Trichopoulos D, Terry KL, Chen WY, Willett WC, Michels KB. A prospective study of breast size and premenopausal breast cancer incidence. *Int J Cancer* 2006; **118**:2031–34.
- Egan KM, Newcomb PA, Titus-Ernstoff L, Trentham-Dietz A, Baron JA, Willett WC. The relation of breast size to breast cancer risk in postmenopausal women (United States). *Cancer Causes Control* 1999; **10**:115–18.
- Tavani A, Pregnolato A, La Vecchia C, Negri E, Favero A, Franceschi S. Breast size and breast cancer risk. *Eur J Cancer Prev* 1996; **5**:337–42.
- 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.
- Coltman CE, Steele JR, McGhee DE. Breast volume is affected by body mass index but not age. *Ergonomics* 2017; **60**:1576–85.
- Lim LY, Ho PJ, Liu J *et al.* Determinants of breast size in Asian women. *Sci Rep* 2018; **8**:1201.
- Brown N, White J, Milligan A *et al.* The relationship between breast size and anthropometric characteristics. *Am J Hum Biol* 2012; **24**:158–64.
- Wade TD, Zhu G, Martin NG. Body mass index and breast size in women: same or different genes? *Twin Res Hum Genet* 2010; **13**:450–54.
- Bhaskaran K, Douglas I, Forbes H, dos-Santos-Silva I, Leon DA, Smeeth L. Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5.24 million UK adults. *Lancet* 2014; **384**:755–65.
- Renahan AG, Tyson M, Egger M, Heller RF, Zwahlen M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet* 2008; **371**:569–78.
- Liu K, Zhang W, Dai Z *et al.* Association between body mass index and breast cancer risk: evidence based on a dose-response meta-analysis. *Cancer Manag Res* 2018; **10**:143–51.
- Schoemaker MJ, Nichols HB, Wright LB *et al.* Association of body mass index and age with subsequent breast cancer risk in premenopausal women. *JAMA Oncol* 2018; **4**:e181771.
- Suzuki R, Orsini N, Saji S, Key TJ, Wolk A. Body weight and incidence of breast cancer defined by estrogen and progesterone receptor status: a meta-analysis. *Int J Cancer* 2009; **124**:698–712.
- Vrieling A, Buck K, Kaaks R, Chang-Claude J. Adult weight gain in relation to breast cancer risk by estrogen and progesterone receptor status: a meta-analysis. *Breast Cancer Res Treat* 2010; **123**:641–49.
- 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.
- Gao C, Patel CJ, Michailidou K *et al.* Mendelian randomization study of adiposity-related traits and risk of breast, ovarian, prostate, lung and colorectal cancer. *Int J Epidemiol* 2016; **45**:896–908.
- Shu X, Wu L, Khankari NK, Shu XO, Wang TJ, Michailidou K, *et al.* Associations of obesity and circulating insulin and glucose with breast cancer risk: a Mendelian randomization analysis. *Int J Epidemiol* 2019; **48**:795–806.
- Bulik-Sullivan B, Finucane HK, Anttila V *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* 2015; **47**:1236–41.
- Hemani G, Zheng J, Elsworth B *et al.* The MR-base platform supports systematic causal inference across the human phenotype. *eLife* 2018; **7**:e34408.
- 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.



23. Pickrell JK, Berisa T, Liu JZ, Ségurel L, Tung JY, Hinds DA. Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet* 2016;**48**:709–17.
24. Michailidou K, Lindström S, Dennis J *et al.* Association analysis identifies 65 new breast cancer risk loci. *Nature* 2017;**551**: 92–94.
25. Michailidou K, Beesley J, Lindstrom S *et al.* Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nat Genet* 2015;**47**: 373–80.
26. Michailidou K, Hall P, Gonzalez-Neira A *et al.* Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet* 2013;**45**:353–61, 61e1–2.
27. Noyce AJ, Kia DA, Hemani G *et al.* Estimating the causal influence of body mass index on risk of Parkinson disease: a Mendelian randomisation study. *PLoS Med* 2017;**14**: e1002314.
28. 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.
29. 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.
30. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* 2013;**37**:658–65.
31. 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.
32. 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.
33. Lindstrom S, Thompson DJ, Paterson AD *et al.* Genome-wide association study identifies multiple loci associated with both mammographic density and breast cancer risk. *Nat Commun* 2014;**5**:5303.
34. Lake JK, Power C, Cole TJ. Women's reproductive health: the role of body mass index in early and adult life. *Int J Obes Relat Metab Disord* 1997;**21**:432–38.
35. Potischman N, Swanson CA, Siiteri P, Hoover RN. Reversal of relation between body mass and endogenous estrogen concentrations with menopausal status. *J Natl Cancer Inst* 1996;**88**: 756–58.
36. Rich-Edwards JW, Goldman MB, Willett WC *et al.* Adolescent body mass index and infertility caused by ovulatory disorder. *Am J Obstet Gynecol* 1994;**171**:171–77.
37. Boyd NF, Guo H, Martin LJ *et al.* Mammographic density and the risk and detection of breast cancer. *N Engl J Med* 2007;**356**: 227–36.
38. Pettersson A, Graff RE, Ursin G *et al.* Mammographic density phenotypes and risk of breast cancer: a meta-analysis. *J Natl Cancer Inst* 2014;**106**. doi: 10.1093/jnci/dju078.
39. Teumer A. Common methods for performing mendelian randomization. *Front Cardiovasc Med* 2018;**5**:51.