Instrumental Heterogeneity in Sex-Specific Two-Sample Mendelian Randomization: Empirical Results From the Relationship Between Anthropometric Traits and Breast/Prostate Cancer

Yixin Gao[1†], Jinhui Zhang[1†], Huashuo Zhao[1,2*], Fengjun Guan[3*] and Ping Zeng[1,2*]

[1] Department of Epidemiology and Biostatistics, School of Public Health, Xuzhou Medical University, Xuzhou, China, [2] Center for Medical Statistics and Data Analysis, School of Public Health, Xuzhou Medical University, Xuzhou, China, [3] Department of Pediatrics, Affiliated Hospital of Xuzhou Medical University, Xuzhou, China

Background: In two-sample Mendelian randomization (MR) studies, sex instrumental heterogeneity is an important problem needed to address carefully, which however is often overlooked and may lead to misleading causal inference.

Methods: We first employed cross-trait linkage disequilibrium score regression (LDSC), Pearson's correlation analysis, and the Cochran's Q test to examine sex genetic similarity and heterogeneity in instrumental variables (IVs) of exposures. Simulation was further performed to explore the influence of sex instrumental heterogeneity on causal effect estimation in sex-specific two-sample MR analyses. Furthermore, we chose breast/prostate cancer as outcome and four anthropometric traits as exposures as an illustrative example to illustrate the importance of taking sex heterogeneity of instruments into account in MR studies.

Results: The simulation definitively demonstrated that sex-combined IVs can lead to biased causal effect estimates in sex-specific two-sample MR studies. In our real applications, both LDSC and Pearson's correlation analyses showed high genetic correlation between sex-combined and sex-specific IVs of the four anthropometric traits, while nearly all the correlation coefficients were larger than zero but less than one. The Cochran's Q test also displayed sex heterogeneity for some instruments. When applying sex-specific instruments, significant discrepancies in the magnitude of estimated causal effects were detected for body mass index (BMI) on breast cancer (P = 1.63E-6), for hip circumference (HIP) on breast cancer (P = 1.25E-20), and for waist circumference (WC) on prostate cancer (P = 0.007) compared with those generated with sex-combined instruments.

Conclusion: Our study reveals that the sex instrumental heterogeneity has non-ignorable impact on sex-specific two-sample MR studies and the causal effects of anthropometric traits on breast/prostate cancer would be biased if sex-combined IVs are incorrectly employed.

Keywords: two-sample Mendelian randomization, sex-specific and sex-combined instrumental variable, sex heterogeneity, causal effect estimation, summary statistics, breast cancer, prostate cancer, anthropometric traits

INTRODUCTION

In the literature of causal inference in observational studies, Mendelian randomization (MR) is a novel statistical method to establish causal relationship between an exposure and an outcome by leveraging genetic variants as instrumental variables (IVs) (Lawlor et al., 2008; Sheehan et al., 2008). To guarantee MR to be valid, each IV needs to satisfy three critical modeling assumptions: (i) strongly associated with the exposure of interest; (ii) not associated with other confounders that are related to both the exposure and the outcome; (iii) influences the outcome only through the pathway of the exposure and does not exhibit any horizontal pleiotropy. The popularity of MR is particularly accelerated by recent successes in large-scale genome-wide association studies (GWASs) (Altshuler et al., 2008; Visscher et al., 2017; McMahon et al., 2019), which make it feasible to choose appropriate single-nucleotide polymorphisms (SNPs) to be eligible instruments for a series of exposures.

However, due to the limitation of data sharing and participant privacy concern, individual-level GWAS datasets are often not accessible; instead, publicly available summary-level statistics are employed in practice, which brings one great benefit that the exposure and the outcome are not required to be measured on the same set of individuals, leading to the so-called two-sample MR study (Lawlor, 2016). Briefly, in the two-sample MR analysis, a genome-wide significant SNP associated with the exposure is first selected as instrument based on which the causal effect is estimated with only marginal effect sizes of the exposure and the outcome. To enhance power, multiple IVs are often leveraged and the individual causal effects can be combined with an inverse-variance weighted (IVW) manner (Burgess et al., 2017). Indeed, the two-sample MR is considerably powerful and flexible and appears technically straightforward to undertake. Due to those reasons, the past few years have witnessed the rapid development and application of MR for causal inference in genetics and epidemiology (Hartwig et al., 2017b; Davies et al., 2018; Zeng and Zhou, 2019a; Yu et al., 2020; Yuan et al., 2020; Liu et al., 2021).

Nevertheless, the two-sample MR still encounters many practical challenges that need to be addressed carefully. For example, the individuals of the two GWASs included in MR studies should take from non-overlapping populations; otherwise, misleading causal effect estimates may be generated (Burgess et al., 2016; Haycock et al., 2016). In addition, because of the difference in SNP effects among diverse populations, the individuals analyzed in the GWASs of exposure and outcome should be of the same ancestry

(Zeng and Zhou, 2019b; Zeng et al., 2019). Besides the two issues mentioned above, another important problem is the sex heterogeneity of instruments arising in two-sample MR studies for sex-specific diseases such as breast cancer or prostate cancer. For instance, it is intuitive and natural to employ female-specific (or male-specific) IVs when evaluating the causal association between exposures and breast cancer (or prostate cancer) *via* the two-sample MR. Here, we do not consider male breast cancer, as it only accounts for less than 1% of cases.

However, this seems not to be true in sex-specific two-sample MR studies in terms of our literature review, and we discover that only a few MR studies mentioned in their analyses the problem of sex-specific instruments (**Supplementary Tables 1, 2**). It is a little surprising that a large amount of sex-specific MR studies exploited sex-combined IVs, which essentially assumed that no sex heterogeneity was present in IVs. However, such an assumption may not hold, since previous GWASs have displayed sex differences in genetic architecture for many exposures including anthropometric traits (**Table 1**). For example, substantial discrepancies were observed at several adiposity-associated loci, and multiple waist-to-hip ratio (WHR)-associated SNPs showed consistently stronger effects in females compared with males (Heid et al., 2011; Randall et al., 2013; Shungin et al., 2015).

Although it is particularly important, a formal investigation about the sex instrumental heterogeneity in sex-specific two-sample MR studies is lacking and its consequence seems to be also overlooked by many MR researchers. As a result, improper causal inference might be generated (Lawlor, 2016; Tan et al., 2019). Therefore, the main goal of this work is to explore the influence of sex instrumental heterogeneity on causal effect estimation in two-sample MR analyses when applying sex-specific and sex-combined effects of IVs. In the following, we first described sex genetic similarity and heterogeneity among instruments and demonstrated that the sex instrumental heterogeneity had a non-ignorable impact on causal inference in the sex-specific two-sample MR; this statement was further supported by our numerical simulation. Moreover, as an illustrative example, we chose breast/prostate cancer as the outcome and four anthropometric traits as exposures to explain the possible consequence in real data analysis. We revealed that the causal effects of anthropometric traits on both breast and prostate cancers would be to some extent changed when using sex-specific IVs compared with those generated with sex-combined instruments. We finally offered several valuable suggestions in practical sex-specific two-sample MR studies.

TABLE 1 | Genetic variants with significant sex difference in effect size for four anthropometric traits.

Traits	Gene	CHR	POS	SNP	Effect		P_Q	Folds	References
					Female	Male			
BMI	ZFP64	20	51,087,862	rs6091540	0.030 (0.004)	0.007 (0.005)	9.05E-05	4.3	Locke et al., 2015
BMI	SEC16B	1	177,889,480	rs543874	0.060 (0.005)	0.034 (0.005)	5.23E-05	1.8	Locke et al., 2015
WHR	LYPLAL1	1	217,820,132	rs2820443	0.062 (0.005)	0.002 (0.005)	2.60E-17	31.0	Shungin et al., 2015
WHR	LYPLAL1	1	217,817,340	rs4846567	0.059 (0.005)	0.005 (0.005)	1.18E-13	11.8	Heid et al., 2011
WHR	GRB14	2	165,221,337	rs10195252	0.054 (0.005)	0.010 (0.005)	1.41E-11	5.4	Heid et al., 2011
WHR	VEGFA	6	43,872,529	rs1358980	0.060 (0.005)	0.015 (0.005)	3.70E-11	4.0	Shungin et al., 2015
WC	OR2W5-NLRP3	1	245,717,763	rs10925060	0.002 (0.005)	0.045 (0.006)	1.70E-08	22.5	Shungin et al., 2015
WC	CCNJL	5	159,626,935	rs17472426	−0.014 (0.009)	0.052 (0.010)	3.90E-08	3.7	Shungin et al., 2015
HIP	KLHL31	6	53,648,294	rs7739232	0.063 (0.011)	−0.004 (0.014)	2.90E-05	15.8	Shungin et al., 2015
HIP	C5-FBXW2	9	122,533,883	rs7044106	0.039 (0.007)	−0.003 (0.008)	1.30E-05	13.0	Shungin et al., 2015
HIP	KLF14	7	130,090,402	rs13241538	0.033 (0.005)	−0.003 (0.005)	2.00E-09	11.0	Shungin et al., 2015

CHR, chromosome; POS, position; SNP, single-nucleotide polymorphism; BMI, body mass index; WHR, waist-to-hip ratio; WC, waist circumference; HIP, hip circumference. Note that the SNP effect of WHR, WC, and HIP was generated under the control of BMI.
The value in the parentheses is the standard error.
P_Q denotes the P value of the Cochran's Q-test. Fold is computed based on the absolute effect and is always shown to be larger than 1.

MATERIALS AND METHODS

Genome-Wide Association Study Dataset Sources and Instrument Selection

We initially obtained sex-combined and sex-specific summary statistics of four anthropometric traits [i.e., body mass index (BMI), waist-to-hip ratio (WHR), waist circumference (WC), and hip circumference (HIP)] for individuals of European ancestry from the Genetic Investigation of ANthropometric Traits (GIANT) Consortium (Locke et al., 2015; Shungin et al., 2015). For each SNP in the GIANT study, the association was performed while adjusting for age, age^2 and study-specific covariates via linear regression. In addition, the SNP effect of WHR, WC, or HIP was estimated under the control of BMI. Based on these GWAS datasets, we yielded a set of uncorrelated associated SNPs ($P < 5.00E-8$) to serve as sex-combined or sex-specific IVs for each anthropometric trait (**Supplementary Table 3**).

We next acquired summary statistics of breast cancer from the Breast Cancer Association Consortium (BCAC) (Michailidou et al., 2017) and summary statistics of prostate cancer from the Prostate Cancer Association Group to Investigate Cancer-Associated Alterations in the Genome (PRACTICAL) consortium (Schumacher et al., 2018). In the GWAS of the two types of cancer, the association was also undertaken with individuals of European descent. The SNP effect size was estimated via logistic regression with principal components as covariates and sometimes additionally adjusted for study-specific covariates. The GWAS datasets employed in our study are summarized in **Table 2**.

Detection of Sex Genetic Similarity and Heterogeneity for Anthropometric Traits

To evaluate genetic similarity between the sex-combined and sex-specific effects of anthropometric traits as well

as between the male-specific and female-specific effects of anthropometric traits, we applied cross-trait linkage disequilibrium score regression (LDSC) to calculate overall

TABLE 2 | Summary information of the GWAS datasets employed in our sex-specific two-sample MR analysis.

Trait	Sample size (case/control)	k_0	M	References
BMI				
Sex-combined	322,154	97	2,517,828	Locke et al., 2015
Female-specific	171,977	38	2,459,695	Locke et al., 2015
Male-specific	152,893	30	2,443,565	Locke et al., 2015
WHR				
Sex-combined	210,086	39	2,542,431	Shungin et al., 2015
Female-specific	116,742	34	2,467,778	Shungin et al., 2015
Male-specific	93,480	3	2,146,136	Shungin et al., 2015
WC				
Sex-combined	231,355	70	2,545,772	Shungin et al., 2015
Female-specific	127,470	25	2,473,035	Shungin et al., 2015
Male-specific	104,079	29	2,294,965	Shungin et al., 2015
HIP				
Sex-combined	211,117	89	2,540,653	Shungin et al., 2015
Female-specific	117,340	41	2,466,814	Shungin et al., 2015
Male-specific	93,965	31	2,188,855	Shungin et al., 2015
Breast cancer Female-specific	228,951 (122,977/105,974)		13,011,123	Michailidou et al., 2017
Prostate cancer Male-specific	140,306 (79,194/61,112)		16,486,833	Schumacher et al., 2018

BMI, body mass index; WHR, waist-to-hip ratio; WC, waist circumference; HIP, hip circumference; SNP, single-nucleotide polymorphism; GWAS, genome-wide association study; MR, Mendelian randomization.
Note that k_0 denotes the number of candidate instruments that were obtained directly with those associated SNPs (P < 5.00E-8) reported in the original GWAS papers; M is the total number of SNPs.

genetic correlation ρ_g with all available SNPs (Bulik-Sullivan et al., 2015). The LD scores were computed with genotypes of 503 European individuals in the 1000 Genomes Project (The 1000 Genomes Project Consortium, 2015) and then regressed on the product of Z statistics of two traits. The regression slope provides an unbiased estimate for ρ_g. The software of LDSC (version v1.0.1) was downloaded from https://github.com/bulik/ldsc.

We also carried out the Pearson's correlation analysis to quantify the genetic effect correlation r_g among those independently associated SNPs that served as IVs. Note that, unlike ρ_g, which quantifies the global genetic overlap of two traits using genome-wide variants, r_g can be viewed as a measurement of marginal genetic sharing between two traits because of only a small set of associated SNPs involved. We further assessed sex heterogeneity in each IV via the Cochran's Q test. Specifically, the sex heterogeneity was tested based on sex-specific SNP effect estimates and standard errors; the P value of heterogeneity was corrected with the Bonferroni's method to take multiple comparisons into account. Finally, the sex heterogeneity was quantified with the I^2 statistic that was widely used in the literature.

Simulation Study to Assess the Influence of Sex Instrumental Heterogeneity

Given the potential sex instrumental heterogeneity in IVs (Heid et al., 2011; Randall et al., 2013; Shungin et al., 2015), we implemented a simple simulation to assess its influence on causal inference in sex-specific two-sample MR studies. To obtain exposures, we first generated m uncorrelated genetic variants with m following a uniform distribution ranging from 50 to 150. The minor allele frequency (MAF) of these SNPs was independently sampled from a uniform distribution ranging from 0.01 to 0.50. The genotypes (denoted by G_1 or G_2) were separately generated for N_1 (1×10^5) male or N_2 (1×10^5) female individuals under the assumption of linkage equilibrium and Hardy–Weinberg equilibrium (HWE). The effect sizes (denoted by α_1 and α_2) for the two sets of SNPs were drawn from a bivariate normal distribution with $\mu = (0, 0)$ and $\Sigma = ([1, r_g], [r_g, 1])$, with r_g varying from 0.1, 0.3, 0.5, to 0.7. Note that r_g also partly quantified the genetic effect heterogeneity of the exposure between females and males for these SNPs; that is, smaller r_g indicated larger heterogeneity. The residual error terms (denoted by e_1 or e_2) were separately sampled from independent standard normal distributions. We further rescaled α_1 and α_2 so that the phenotypic variance explained (PVE) by SNPs can be set to the given value for the male exposure and the female exposure.

$$\text{PVE}_l = \frac{\delta_l^2 \times \sum_{j=1}^{m} \alpha_{lj}^2}{\delta_l^2 \times \sum_{j=1}^{m} \alpha_{lj}^2 + 1}, l = 1 \text{ or } 2 \tag{1}$$

where δ is the scale parameter and can be estimated (denoted by $\hat{\delta}$) in terms of (1). Note that r_g would not be impacted by the scaled effect sizes. After doing

those, the exposures (denoted by x_1 and x_2) can be obtained.

$$x_{1i} = \hat{\delta}_1 \times \sum_{j=1}^{K} G_{1ij}\alpha_{1j} + e_{1i}, x_{2i} = \hat{\delta}_2 \times \sum_{j=1}^{K} G_{2ij}\alpha_{2j}$$
$$+ e_{2i}, i = 1, \cdots, N_1 \text{ or } N_2 \tag{2}$$

Then, a *female-specific* outcome was created as $y = x_2 \times \theta + \varepsilon$ based on the same set of individuals, where θ was the true causal effect size varying from 0.1, 0.3, to 0.5 and ε was the residual error term following a standard normal distribution. The single SNP association analysis was conducted to obtain summary statistics for each genetic variant (Zeng et al., 2015). Specifically, the sex-specific summary statistics of the exposure were yielded by regressing x_1 on G_1 (or x_2 on G_2), while the sex-combined summary statistics of the exposure were yielded with the fixed-effects IVW meta-analysis based on the two sets of sex-specific summary statistics. The SNPs with marginal P values less than $0.05/m$ were identified to be female-specific or sex-combined IVs. If no SNPs satisfied this criterion, the one with the minimum P value would be employed. With the selected IVs, the female-specific and sex-combined causal effects of the exposure on the outcome were evaluated with the fixed-effects IVW MR approach (Burgess et al., 2017; Hartwig et al., 2017a; Yavorska and Burgess, 2017).

$$\hat{\theta} = \frac{1}{\sum_{i=1}^{k} \text{var}(\hat{a}_i^y)^{-1}(\hat{a}_i^x)^2} \sum_{i=1}^{k} \text{var}(\hat{a}_i^y)^{-1}\hat{a}_i^y\hat{a}_i^x \tag{3}$$

where $\hat{a}$ denotes the marginal effect size of the selected IV, $\hat{a}$ denotes the corresponding variance of the estimated effect size, and k is the number of the used IVs. Note that $\hat{\theta} = \hat{a}^y/\hat{a}^x$ when only one IV was employed. We can approximately estimate the asymptotic variance of $\hat{\theta}$ by:

$$\text{var}(\hat{\theta}) = \frac{1}{\sum_{i=1}^{k} \text{var}(\hat{\alpha}_i^y)^{-1}(\hat{\alpha}_i^x)^2} \tag{4}$$

Estimation of Causal Effect With Two-Sample Mendelian Randomization

Before the formal MR analysis, we performed several stringent quality control procedures for IVs: (i) excluded SNPs that were not included in the breast/prostate cancer GWASs; (ii) excluded SNPs whose alleles were inconsistent between the exposure and the outcome; (iii) excluded SNPs that were likely related to breast/prostate cancer if the Bonferroni-corrected P values < 0.05. Note that this is a conservative way of protecting against the pleiotropic impact of instruments to ensure valid causal inference in MR studies (Zeng and Zhou, 2019a; Zeng et al., 2019). After the quality control, we performed the two-sample IVW MR to estimate the causal effect of each anthropometric trait on breast/prostate cancer with sex-combined or sex-specific IVs. The causal relationship was mainly illustrated in terms of odds ratio (OR) per standard

deviation (SD) increase in anthropometric trait because all the anthropometric traits were previously standardized.

For sex-specific IVs, once a significant causal relationship was identified between one of the anthropometric traits and breast/prostate cancer *via* the IVW approach, we further implemented the weighted median method (Bowden et al., 2016b), the maximum likelihood method (Burgess et al., 2013), the leave-one-out (LOO) analysis (Noyce et al., 2017), the MR-PRESSO test (Verbanck et al., 2018), and the MR-Egger regression (Bowden et al., 2016a; Burgess and Thompson, 2017) as part of sensitivity analyses to examine the robustness of our results.

Finally, we formally compared the causal effects generated with sex-combined or sex-specific instruments (H_0: $\theta_{combined} = \theta_{specific}$) by a simple u test:

$$u = \frac{\hat{\theta}_{combined} - \hat{\theta}_{specific}}{\sqrt{\begin{array}{c} \{se(\hat{\theta}_{combined})\}^2 + \{se(\hat{\theta}_{specific})\}^2 \\ - 2\rho \times se(\hat{\theta}_{combined}) \times se(\hat{\theta}_{specific}) \end{array}}} \quad (5)$$
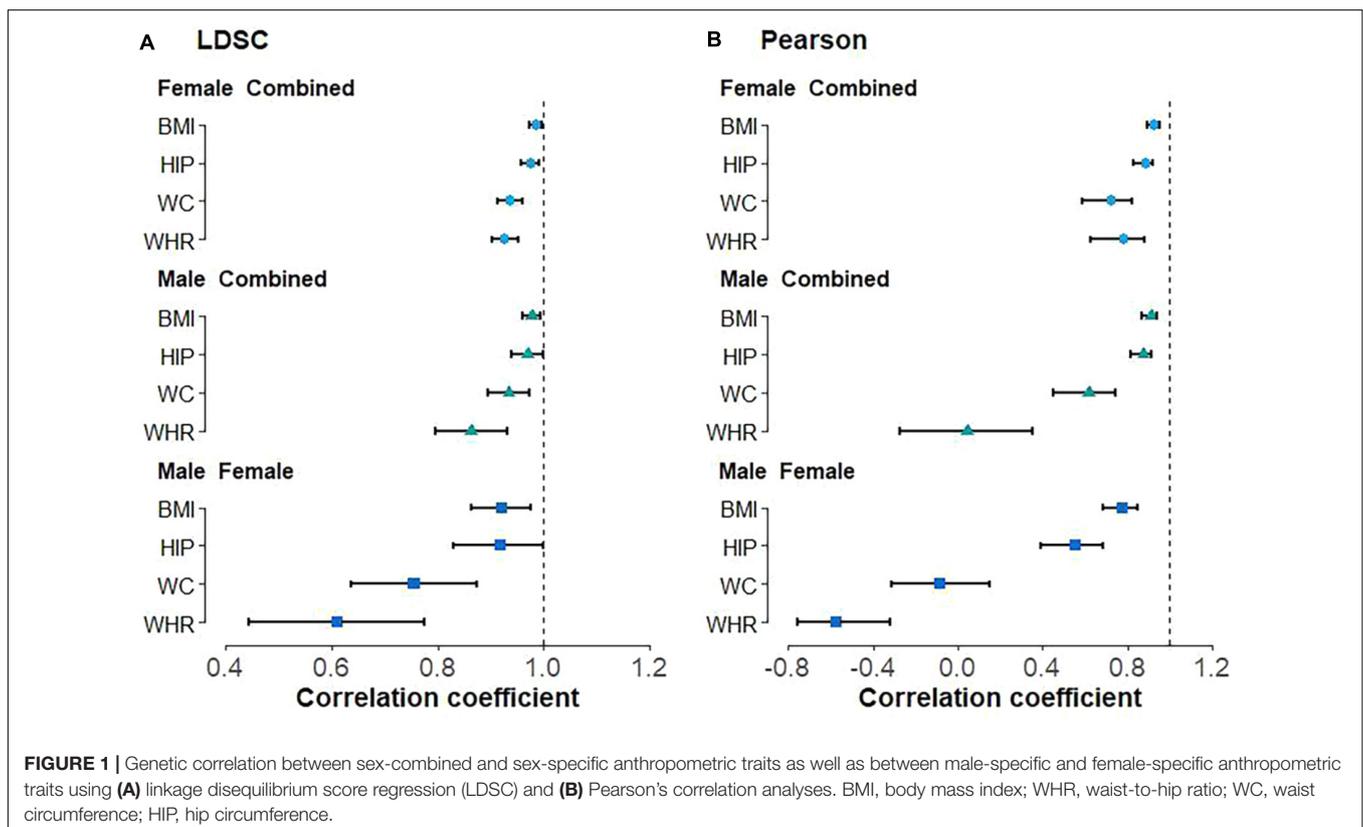
where $\hat{\theta}$ is the estimated causal effect with the standard error $se(\hat{\theta})$, and ρ is the correlation coefficient between the two estimated causal effects that are expected to be highly correlated. However, it is rather challenging to estimate ρ accurately with only summary statistics. In the present study, we calculated ρ with the LOO jackknife method for each anthropometric trait (Efron, 1982; Efron and Tibshirani, 1993). Specifically, for all

the used sex-combined and sex-specific IVs (**Supplementary Table 3**), we removed one at a time and recomputed the causal effect on breast/prostate cancer using the rest of sex-combined and sex-specific instruments, respectively. We then calculated ρ using the sex-combined causal effects and the sex-specific causal effects. The P value of the u statistic can be easily calculated as it asymptotically follows a standard normal distribution.

RESULTS

Sex Heterogeneity of Instrumental Variables

With LDSC, we observe high positive genetic correlation between male and female anthropometric traits (**Figure 1A**), with $\hat{\rho}_g$ = 0.920 (se = 0.028) for BMI, $\hat{\rho}_g$ = 0.919 (se = 0.046) for HIP, $\hat{\rho}_g$ = 0.755 (se = 0.060) for WC, and $\hat{\rho}_g$ = 0.609 (se = 0.085) for WHR. Similar positive genetic correlations were also detected between the sex-combined and sex-specific anthropometric traits, with all $\hat{\rho}_g$ larger than 0.85 (**Figure 1A**). Nearly all these estimated genetic correlations are substantially larger than 0 (H_{01}: $\rho_g = 0$) but less than 1 (H_{02}: $\rho_g = 1$) (**Figure 1A**); the only exception is the genetic correlation of HIP between males and females, which is marginally significant ($\hat{\rho}_g$ = 0.919 and $P = 0.076$ for H_{02}). In addition, the Pearson's correlation analysis also showed that all the correlations between the sex-combined and sex-specific instruments are less than 1 (H_{30}: $r_g = 1$) (**Figure 1B**, **Supplementary Figure 1**). It needs to note that two negative

FIGURE 1 | Genetic correlation between sex-combined and sex-specific anthropometric traits as well as between male-specific and female-specific anthropometric traits using **(A)** linkage disequilibrium score regression (LDSC) and **(B)** Pearson's correlation analyses. BMI, body mass index; WHR, waist-to-hip ratio; WC, waist circumference; HIP, hip circumference.

Pearson's correlations even occur, i.e., $\hat{r}_g = -0.087$ ($P = 0.472$) for WC and $\hat{r}_g = -0.577$ ($P = 1.69\text{E-}5$) for WHR between males and females (**Supplementary Figures 1C3, D3**), indicating that the marginal genetic effect correlation can have a fully opposite direction compared with its global counterpart.

Using the Cochran's Q test, we detect two out of 97 (2.1%) BMI-associated SNPs, four of 95 (4.2%) HIP-associated SNPs, 10 of 76 (13.2%) WC-associated SNPs, and 17 of 48 (35.4%) WHR-associated SNPs exhibit sex heterogeneity among candidate IVs (**Supplementary Table 3**). In particular, all WC-associated SNPs but one (i.e., rs17472426 with $\beta_{male} = 0.052$ and $\beta_{female} = -0.014$, $P_{het} = 9.31\text{E-}07$) and all WHR-associated SNPs but three (i.e., rs224333 with $\beta_{male} = 0.036$ and $\beta_{female} = 0.009$, $P_{het} = 1.34\text{E-}4$; rs10925060 with $\beta_{male} = 0.045$ and $\beta_{female} = 0.002$, $P_{het} = 3.68\text{E-}8$; rs3791679 with $\beta_{male} = 0.053$ and $\beta_{female} = 0.021$, $P_{het} = 4.18\text{E-}5$) are found to have larger effect sizes on females compared with males (**Supplementary Table 3**). These findings, together with the estimates of genetic correlation described above, indicate the existence of potential sex genetic effect heterogeneity in the four anthropometric traits, although they indeed share widely common genetic components.

Influence of Sex Instrumental Heterogeneity in Terms of the Simulation

The results of simulation are displayed in **Figure 2**. We here clearly find that the use of sex-combined instruments can lead to biased causal effect estimates in our simulated case of *female-specific* two-sample MR. Specifically, the female-specific causal effect is overestimated if the PVE of the male exposure is smaller than that of the female exposure (1% vs. 3%) (**Figures 2A–C**). The main reason is that under this situation, the average of the sex-combined effect sizes of the IVs tends to be smaller compared to the average of female-specific effects of the IVs [see Eq. 1], which in turn leads to a higher estimate of causal effect when the effect sizes of the IVs of the outcome remain fixed [see Eq. 3]. On the other hand, the female-specific causal effect is underestimated if the PVE of the male exposure is larger than that of the female exposure (3% vs. 1%) (**Figures 2D–F**).

Moreover, as can be expected, the sex-combined bias generally becomes larger as the sex heterogeneity in IVs increases (i.e., weaker correlation). For example, when true effect size is 0.5 (**Figure 2C**), the average of the bias is 0.169 for $r_g = 0.1$, 0.150 for $r_g = 0.3$, 0.131 for $r_g = 0.5$, or 0.127 for $r_g = 0.7$. In contrast, approximate unbiased estimates of causal effects are generated when using sex-specific IVs. Overall, this simulation study explicitly reveals that the use of sex-combined IVs in sex-specific two-sample MR analyses may lead to a biased causal effect estimate in many cases and that the extension of the bias relies on the relative magnitude of the sex-specific effects and the sex-combined effects of the used IVs.
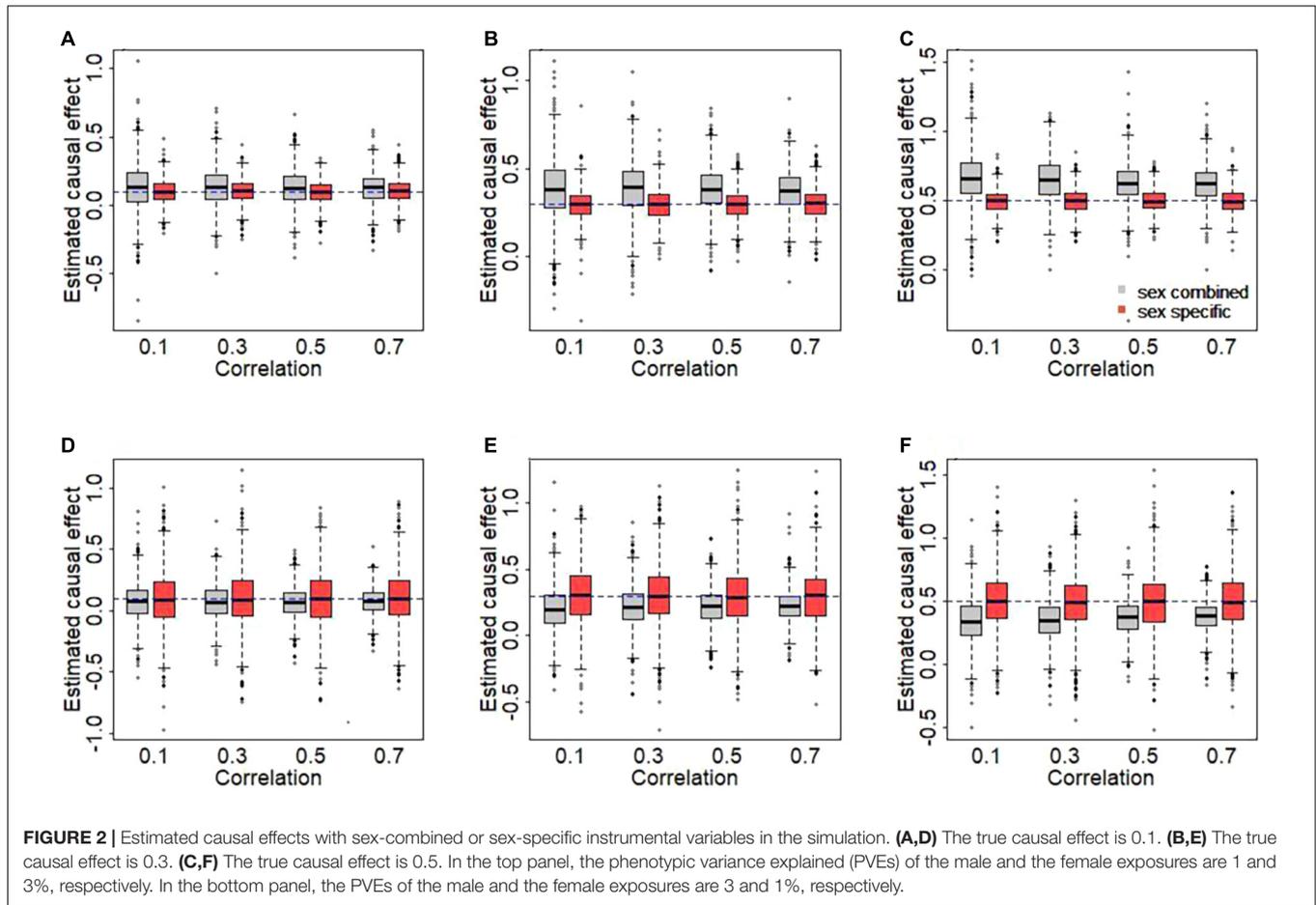
Causal Effect Estimation With Sex-Combined Instruments

Like most previous studies (**Supplementary Tables 1, 2**), we first employ sex-combined instruments to evaluate the causal relationship between each anthropometric trait and breast/prostate cancer as an exploratory analysis. With the random-effects IVW method, which can account for instrumental heterogeneity, we identify statistically significant associations between BMI and breast cancer [OR = 0.85, 95% confidence interval (CIs) 0.76∼0.95, $P = 0.003$], between WC and breast cancer (OR = 0.87, 95% CI 0.77∼0.98, $P = 0.020$), as well as between BMI and prostate cancer (OR = 0.87, 95% CI 0.76∼0.98, $P = 0.022$) (**Tables 3, 4**), indicating that higher BMI can lead to a lower risk for breast/prostate cancer and that higher WC can result in reduced risk for breast cancer. These results are considerably consistent with prior observations (Shu et al., 2018; Kazmi et al., 2019; Qian et al., 2019). After the removal of IVs exhibiting sex heterogeneity in terms of the Cochran's Q test (**Supplementary Tables 4, 5**), we observe that the association between WC and breast cancer now becomes non-significant ($P = 0.069$), although other causal effects almost remain unchanged because only a few instruments are excluded, indicating that the sex heterogeneity in IVs might have a substantial influence on the statistical inference in sex-specific two-sample MR analyses.

Causal Effect Estimation With Sex-Specific Instruments

We now estimate the causal effect using only sex-specific instruments and show those new results in **Tables 3, 4**. Several interesting findings are observed. First, it is observed that there exists a significantly positive correlation between the causal effect obtained with sex-specific effects of IVs and that yielded with sex-combined effects (**Supplementary Figures 2, 3**), with all the correlation coefficients [i.e., ρ in the u test in Eq. 5] larger than 0.9. Second, using the sex-specific IVs, we discover distinct discrepancy in the significance of estimated causal effects of anthropometric traits compared with those obtained with the sex-combined IVs. Specifically, although still maintaining similar effect sizes (0.866 vs. 0.858), WC is now only marginally associated with breast cancer ($P = 0.057$) (**Table 3**). The analogous situation is seen for the association between BMI and prostate cancer, which also becomes marginally significant ($P = 0.060$), although the causal effect is not influenced (0.865 vs. 0.862) (**Table 4**). Third, when applying sex-specific IVs, significant discrepancies in the magnitude of estimated causal effects are detected for BMI on breast cancer ($P = 1.63\text{E-}6$), for HIP on breast cancer ($P = 1.25\text{E-}20$), as well as for WC on prostate cancer ($P = 0.007$) compared with those estimated with sex-combined instruments (**Tables 3, 4**). More specifically, we find that the causal association between BMI and breast cancer is now more pronounced (**Table 3**). For example, it is shown that per SD increase of BMI can result in about 23.7% (95% CI 11.8%∼33.9%) lower risk of breast cancer when using the female-specific instruments (OR = 0.76) compared with 15.5% (95% CI 5.5%∼24.4%) lower risk of breast cancer if using those sex-combined instruments (OR = 0.85). Here, it needs to be highlighted that the female PVE of BMI is much larger than the male PVE of BMI (1.16 vs. 0.77; see **Tables 3, 4**), partly explaining why a stronger association was identified when the female-specific effects of instruments were employed. Note that

FIGURE 2 | Estimated causal effects with sex-combined or sex-specific instrumental variables in the simulation. **(A,D)** The true causal effect is 0.1. **(B,E)** The true causal effect is 0.3. **(C,F)** The true causal effect is 0.5. In the top panel, the phenotypic variance explained (PVEs) of the male and the female exposures are 1 and 3%, respectively. In the bottom panel, the PVEs of the male and the female exposures are 3 and 1%, respectively.

TABLE 3 | Association of anthropometric trait with the risk of breast cancer using sex-combined and female-specific instruments.

Exposure	Sex-combined IVs				Female-specific IVs				u (P)
	k_0/k_1	PVE (%)	Power (%)	OR (95% CI, P)	k_0/k_1	PVE (%)	Power (%)	OR (95% CI, P)	
BMI	97/92	1.51	99.9	0.845 (0.756∼0.945, 0.003)	38/36	1.16	100.0	0.763 (0.661∼0.882, 2.52E-04)	4.795 (1.63E-06)
WHR	39/38	0.95	2.8	1.002 (0.841∼1.194, 0.984)	34/34	1.63	6.7	0.985 (0.858∼1.131, 0.833)	0.610 (0.542)
WC	70/66	1.42	98.0	0.866 (0.767∼0.978, 0.020)	25/25	1.04	96.1	0.858 (0.732∼1.004, 0.057)	0.299 (0.765)
HIP	89/86	2.07	6.1	0.988 (0.885∼1.102, 0.823)	41/39	1.57	93.6	0.890 (0.787∼1.006, 0.063)	9.669 (1.25E-20)

IV, instrumental variable; BMI, body mass index; WHR, waist-to-hip ratio; WC, waist circumference; HIP, hip circumference; MR, Mendelian randomization; PVE, phenotypic variance explained by IVs that were used in our final MR analysis; power, statistical power; OR, odds ratio; CI, confidence internal.
Note that k_0 is the number of candidate instruments, while k_1 is the final number of instruments employed in the analysis.

TABLE 4 | Association of anthropometric trait with the risk of prostate cancer using sex-combined and male-specific instruments.

Exposure	Sex-combined IVs				Male-specific IVs				u (P)
	k_0/k_1	PVE (%)	Power (%)	OR (95% CI, P)	k_0/k_1	PVE (%)	Power (%)	OR (95% CI, P)	
BMI	97/60	1.12	81.3	0.865 (0.764∼0.979, 0.022)	30/22	0.77	67.7	0.862 (0.738∼1.007, 0.060)	0.190 (0.849)
WHR	39/21	0.56	59.2	0.854 (0.681∼1.071, 0.172)	3/1	0.04	5.2	1.094 (0.545∼2.198, 0.800)	−0.969 (0.333)
WC	70/50	1.02	14.1	0.954 (0.808∼1.126, 0.577)	29/21	1.15	59.0	0.895 (0.763∼1.050, 0.175)	2.688 (0.007)
HIP	89/64	1.40	5.4	1.016 (0.886∼1.166, 0.817)	31/17	0.83	28.6	1.086 (0.873∼1.350, 0.457)	−1.428 (0.153)

IV, instrumental variable; BMI, body mass index; WHR, waist-to-hip ratio; WC, waist circumference; HIP, hip circumference; MR, Mendelian randomization; PVE, phenotypic variance explained by IVs that were used in our final MR analysis; power, statistical power; OR, odds ratio; CI, confidence internal.
Note that k_0 is the number of candidate instruments, while k_1 is the final number of instruments employed in the analysis.

this finding is also consistent with the phenomenon discovered in the simulation above.

Sensitivity Analyses With Sex-Specific Instruments

We here perform a wide series of sensitivity analyses to complement our main MR results obtained above using IVW with sex-specific IVs. Here, only the significant relationship between BMI and breast cancer is considered (**Table 3**). Both the weighted median method and the maximum likelihood method yield consistent causal effect estimates compared with IVW (**Figure 3**). We also create a scatter plot to examine whether instrumental outliers are present. Among all these used female-specific instruments, one index SNP (i.e., rs17024393 on gene *GNAT2*) has a large effect size of 0.071 on BMI and may be likely a potential outlier (**Supplementary Figure 4A**). However, after removing this IV, the OR is estimated to be 0.78 (95% CI 0.67∼0.90, P = 5.28E-04) (**Supplementary Figure 4C**), almost the same as that obtained using all instruments together (OR = 0.76). In addition, the LOO analysis shows that no single instrument can substantially influence the overall IVW estimate (**Supplementary Figure 4C**). The result of MR-PRESSO displays the absence of instrument outliers at the significance level of 0.05. Finally, the intercept of MR-Egger regression is estimated as 0.006 (95% CI −0.009∼0.022, P = 0.418), ruling out the possibility of directional pleiotropic effects of instruments. The funnel plot also presents a symmetric pattern around the overall point estimate (**Supplementary Figure 4B**), implying that horizontal pleiotropy unlikely biases our result. In conclusion, depending on female-specific IVs, we demonstrate that BMI is robustly negatively associated with breast cancer.

Power Calculation

Finally, we calculate the statistical power to detect the estimated causal effects of four anthropometric traits on the risk of breast or prostate cancer when applying sex-specific and sex-combined IVs. The power calculation is based on observed causal effect sizes, the number of IVs used, the sample sizes of the exposure, and the outcome. We implement the power calculation via an online software tool available at https://sb452.shinyapps.io/power (Brion et al., 2013).
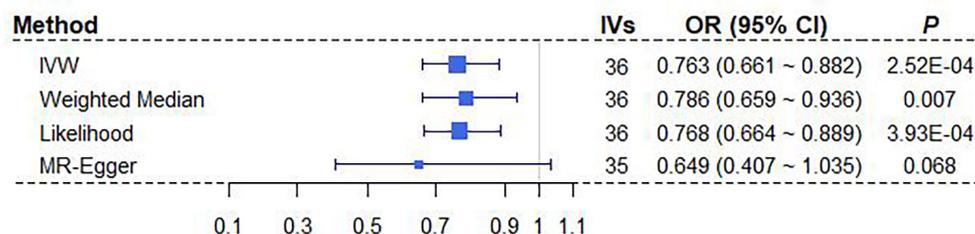
With regard to the association between WC and breast cancer, the statistical power calculated with female-specific instruments

(power = 96.1% when IVs = 25 and PVE = 1.04%) is slightly lower than that computed with sex-combined instruments (power = 98.3% when IVs = 66 and PVE = 1.42%) (**Table 3**). The analogous situation is seen for the association between BMI and prostate cancer (**Table 4**). In these analyses, we observe that the number of valid IVs would become smaller after applying sex-specific instruments, which can potentially reduce the statistical power in the MR analysis due to the decrease of PVE.

However, it is also shown that although there is a decrease in the number of valid IVs after using sex-specific instruments, the statistical powers of some associations are higher than those obtained with sex-combined ones (**Tables 3, 4**). For example, an improvement of statistical power is detected for BMI on breast cancer (power = 100.0% when IVs = 36 vs. power = 99.9% when IVs = 92), for HIP on breast cancer (power = 93.6% when IVs = 39 vs. power = 6.1% when IVs = 86), as well as for WC on prostate cancer (power = 59.9% when IVs = 21 vs. power = 14.1% when IVs = 50) after applying sex-specific instruments. As mentioned before, these analyses show significant discrepancies in the magnitude of estimated causal effects. Therefore, we cannot completely rule out the possibility that the difference in the causal inference is due to the different number of IVs used. However, as shown in our real applications, the impact of sex instrumental heterogeneity on the magnitude of estimated causal effects is substantial, which is not fully driven by the decreased number of IVs.

DISCUSSION

The main objective of our study was to investigate the influence of sex instrumental heterogeneity on causal effect estimation in sex-specific two-sample MR analyses where in principle sex-specific rather than sex-combined IVs should be employed. One of the common cases is the application of MR to breast cancer or prostate cancer, both of which lead to serious health threat in females and males worldwide (Bray et al., 2018). Therefore, the identification of risk factors for the two types of cancer is important for disease prevention and holds the potential for better therapeutic strategies in the future. More than 80 empirical MR studies have been implemented for the two cancers to date (**Supplementary Tables 1, 2**); however, only few applied sex-specific IVs for exposures in these analyses (Au Yeung and Schooling, 2019; Jung et al., 2019; Ooi et al., 2019; Li et al., 2020;

Method	IVs	OR (95% CI)	P
IVW	36	0.763 (0.661 ∼ 0.882)	2.52E-04
Weighted Median	36	0.786 (0.659 ∼ 0.936)	0.007
Likelihood	36	0.768 (0.664 ∼ 0.889)	3.93E-04
MR-Egger	35	0.649 (0.407 ∼ 1.035)	0.068

FIGURE 3 | Estimated causal effects of BMI on breast cancer with different MR approaches. MR-Egger regression was performed after removing the instrument of rs17024393. BMI, body mass index; IV, instrumental variable; IVW, inverse-variance weighted; OR, odds ratio; MR, Mendelian randomization.

Mohammadi-Shemirani et al., 2020; Murphy et al., 2020; Ong et al., 2020; Richardson et al., 2020; Ruth et al., 2020; Watts et al., 2020). Particularly, for some exposures (e.g., WHR) that display obvious sex heterogeneity in effect sizes, sex-combined IVs were still employed; even sex-specific summary statistics were publicly available (Gao et al., 2016; Shu et al., 2018). Here, we emphasize again that the issue of sex instrumental heterogeneity specially occurs only in summary statistics-based MR studies, which does not arise when individual-level datasets can be accessible.

In our simulation and empirical analyses, we illustrated that the sex instrumental heterogeneity had a non-ignorable impact on the causal inference and that the causal effects of anthropometric traits on breast/prostate cancer would be greatly influenced if sex-combined IVs were incorrectly applied. To the best of our knowledge, our study is among the first to formally examine the problem of sex-specific instruments in the two-sample MR.

In sex-specific two-sample MR studies, the use of sex-combined instruments makes an implicit hypothesis that no effect differences exist between females and males, which, however, is not always satisfied in terms of previous observations (e.g., **Table 1**). Nevertheless, in practice, applying sex-combined instruments in MR is not without advantages if such assumption can be well-established. Under this situation, one of the greatest benefits is that more IVs would be exploited because of a larger sample size for the exposure GWAS, which can potentially lead to the improvement of statistical power due to more phenotypic variances explained (e.g., **Tables 3**, **4**).

From a more generalized perspective, the females and the males can be viewed as two diverse populations that have different genetic foundations (Ober et al., 2008) as well as distinct morbidities and mortalities of complex diseases (Wang et al., 2019). Note that to ensure the validity of two-sample MR, one of the important assumptions is that two sample sets should come from the same underlying population. Otherwise, MR may still provide evidence on whether a causal association exists but not necessarily on the precise magnitude of the causal effect (Burgess et al., 2015; Haycock et al., 2016). Our study demonstrated that sex-specific instruments can substantially influence the significance and magnitude of causal effects, confirming the importance of this MR assumption. At the same time, we note that a few sex-specific two-sample MR studies applying sex-specific IVs were published in recent years (**Table 1**), suggesting the growing attention has been paid on the issue of sex instrumental heterogeneity.

Finally, we offer some suggestions for the issue of sex heterogeneity of instruments when conducting sex-specific two-sample MR studies. First, to guarantee to implement MR in the same population, it is necessary to check the original GWAS of exposures or contact authors to obtain sex-specific summary statistics; a clear statement about whether sex-specific IVs are employed is also highly recommended (Ong et al., 2020; Richardson et al., 2020). Second, when sex-combined instruments are employed, sex heterogeneity in instruments should be carefully examined, followed by extensive sensitivity analyses. When using sex-specific instruments, various sensitivity analyses also should be carried out to ensure the robustness of the results. Third, when only sex-combined summary statistics are available and one has to apply sex-combined IVs for exposures, explaining possible biases in the causal inference due to sex instrumental heterogeneity is strongly advocated (Au Yeung and Schooling, 2019; Jiang et al., 2020; Papadimitriou et al., 2021).

In summary, although the two-sample MR is technically easy to undertake, the principal modeling assumptions should still be validated in sex-specific two-sample MR studies strictly. Especially in the sex-specific two-sample MR analyses, the choice of appropriate IVs for exposures can reduce the bias of causal effect estimation and make MR results more reliable.

CONCLUSION

Our study reveals that the sex instrumental heterogeneity may have a non-ignorable impact on sex-specific two-sample MR studies, and the causal effects of anthropometric traits on breast/prostate cancer would be biased if sex-combined IVs are incorrectly employed.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

PZ conceived the idea for the study. PZ, FG, and HZ obtained the data. PZ and YG cleared up the datasets, performed the data analyses, and drafted the manuscript. PZ, FG, HZ, YG, and JZ interpreted the results of the data analyses. All authors approved the manuscript and provided relevant suggestions.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.651332/full#supplementary-material

REFERENCES

Altshuler, D., Daly, M., and Lander, E. (2008). Genetic mapping in human disease. *Science* 322, 881–888.

Au Yeung, S. L., and Schooling, C. M. (2019). Impact of glycemic traits, type 2 diabetes and metformin use on breast and prostate cancer risk: a Mendelian randomization study. *BMJ Open Diabetes Res. Care* 7:e000872. doi: 10.1136/bmjdrc-2019-000872

Bowden, J., Del Greco, M. F., Minelli, C., Smith, G. D., Sheehan, N. A., and Thompson, J. R. (2016a). Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I-2 statistic. *Int. J. Epidemiol.* 45, 1961–1974. doi: 10.1093/ije/dyw220

Bowden, J., Smith, G. D., Haycock, P. C., and Burgess, S. (2016b). Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. *Genetic Epidemiol.* 40, 304–314. doi: 10.1002/gepi.21965

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 68, 394–424. doi: 10.3322/caac.21492

Brion, M. J., Shakhbazov, K., and Visscher, P. M. (2013). Calculating statistical power in Mendelian randomization studies. *Int. J. Epidemiol.* 42, 1497–1501. doi: 10.1093/ije/dyt179

Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P. R., et al. (2015). An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* 47, 1236–1241. doi: 10.1038/ng.3406

Burgess, S., Butterworth, A., and Thompson, S. G. (2013). Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* 37, 658–665. doi: 10.1002/gepi.21758

Burgess, S., Davies, N. M., and Thompson, S. G. (2016). Bias due to participant overlap in two-sample Mendelian randomization. *Genetic Epidemiol.* 40, 597–608. doi: 10.1002/gepi.21998

Burgess, S., Scott, R. A., Timpson, N. J., Davey Smith, G., and Thompson, S. G. (2015). Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. *Eur. J. Epidemiol.* 30, 543–552. doi: 10.1007/s10654-015-0011-z

Burgess, S., Small, D. S., and Thompson, S. G. (2017). A review of instrumental variable estimators for Mendelian randomization. *Stat. Methods Med. Res.* 26, 2333–2355. doi: 10.1177/0962280215597579

Burgess, S., and Thompson, S. G. (2017). Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur. J. Epidemiol.* 32, 377–389. doi: 10.1007/s10654-017-0255-x

Davies, N. M., Holmes, M. V., and Davey Smith, G. (2018). Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *Br. Med. J.* 362:k601. doi: 10.1136/bmj.k601

Efron, B. (1982). *The Jackknife, The Bootstrap and Other Resampling Plans.* Philadelphia, PA: Society for industrial and applied mathematics.

Efron, B., and Tibshirani, R. (1993). *An Introduction to the Bootstrap.* New York, NY: Chapman and Hall.

Gao, C., Patel, C. J., Michailidou, K., Peters, U., Gong, J., Schildkraut, J., et al. (2016). Mendelian randomization study of adiposity-related traits and risk of breast, ovarian, prostate, lung and colorectal cancer. *Int. J. Epidemiol.* 45, 896–908. doi: 10.1093/ije/dyw129

Hartwig, F. P., Davey Smith, G., and Bowden, J. (2017a). Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int. J. Epidemiol.* 46, 1985–1998. doi: 10.1093/ije/dyx102

Hartwig, F. P., Davies, N. M., Hemani, G., and Davey Smith, G. (2017b). Two-sample Mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique. *Int. J. Epidemiol.* 45, 1717–1726. doi: 10.1093/ije/dyx028

Haycock, P. C., Burgess, S., Wade, K. H., Bowden, J., Relton, C., and Davey Smith, G. (2016). Best (but oft-forgotten) practices: the design, analysis, and interpretation of Mendelian randomization studies. *Am. J. Clin. Nutr.* 103, 965–978. doi: 10.3945/ajcn.115.118216

Heid, I. M., Jackson, A. U., Randall, J. C., Winkler, T. W., Qi, L., Steinthorsdottir, V., et al. (2011). Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat. Genet.* 43:1164. doi: 10.1038/ng1111-1164a

Jiang, X., Dimou, N. L., Zhu, Z., Bonilla, C., Lewis, S. J., Lindstrom, S., et al. (2020). Allergy, asthma, and the risk of breast and prostate cancer: a Mendelian randomization study. *Cancer Causes Control* 31, 273–282. doi: 10.1007/s10552-020-01271-7

Jung, S. Y., Mancuso, N., Papp, J., Sobel, E., and Zhang, Z. F. (2019). Post genome-wide gene-environment interaction study: the effect of genetically driven insulin resistance on breast cancer risk using Mendelian randomization. *PLoS One* 14:e0218917. doi: 10.1371/journal.pone.0218917

Kazmi, N., Haycock, P., Tsilidis, K., Lynch, B. M., Truong, T., Martin, R. M., et al. (2019). Appraising causal relationships of dietary, nutritional and physical-activity exposures with overall and aggressive prostate cancer: two-sample Mendelian-randomization study based on 79 148 prostate-cancer cases and 61 106 controls. *Int. J. Epidemiol.* 49, 587–596. doi: 10.1093/ije/dyz235

Lawlor, D. A. (2016). Commentary: two-sample Mendelian randomization: opportunities and challenges. *Int. J. Epidemiol.* 45, 908–915. doi: 10.1093/ije/dyw127

Lawlor, D. A., Harbord, R. M., Sterne, J. A., Timpson, N., and Davey Smith, G. (2008). Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat. Med.* 27, 1133–1163. doi: 10.1002/sim.3034

Li, S., Xu, Y., Zhang, Y., Nie, L., Ma, Z., Ma, L., et al. (2020). Mendelian randomization analyses of genetically predicted circulating levels of cytokines with risk of breast cancer. *NPJ Precis Oncol.* 4:25. doi: 10.1038/s41698-020-00131-6

Liu, L., Zeng, P., Xue, F., Yuan, Z., and Zhou, X. (2021). Multi-trait transcriptome-wide association studies with probabilistic Mendelian randomization. *Am. J. Hum. Genet.* 108, 240–256. doi: 10.1016/j.ajhg.2020.12.006

Locke, A. E., Kahali, B., Berndt, S. I., Justice, A. E., Pers, T. H., Felix, R., et al. (2015). Genetic studies of body mass index yield new insights for obesity biology. *Nature* 518, 197–206. doi: 10.1038/nature14177

McMahon, A., Malangone, C., Suveges, D., Sollis, E., Cunningham, F., Riat, H. S., et al. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 47, D1005–D1012. doi: 10.1093/nar/gky1120

Michailidou, K., Lindstrom, S., Dennis, J., Beesley, J., Hui, S., Kar, S., et al. (2017). Association analysis identifies 65 new breast cancer risk loci. *Nature* 551, 92–94. doi: 10.1038/nature24284

Mohammadi-Shemirani, P., Chong, M., Pigeyre, M., Morton, R. W., Gerstein, H. C., and Paré, G. (2020). Effects of lifelong testosterone exposure on health

and disease using Mendelian randomization. *Elife* 9:e58914. doi: 10.7554/eLife.58914

Murphy, N., Knuppel, A., Papadimitriou, N., Martin, R. M., Tsilidis, K. K., Smith-Byrne, K., et al. (2020). Insulin-like growth factor-1, insulin-like growth factor-binding protein-3, and breast cancer risk: observational and Mendelian randomization analyses with approximately 430 000 women. *Ann. Oncol.* 31, 641–649. doi: 10.1016/j.annonc.2020.01.066

Noyce, A. J., Kia, D. A., Hemani, G., Nicolas, A., Price, T. R., De Pablo-Fernandez, E., et al. (2017). Estimating the causal influence of body mass index on risk of Parkinson disease: a Mendelian randomisation study. *PLoS Med.* 14:e1002314. doi: 10.1371/journal.pmed.1002314

Ober, C., Loisel, D. A., and Gilad, Y. (2008). Sex-specific genetic architecture of human disease. *Nat. Rev. Genet.* 9, 911–922.

Ong, J. S., Derks, E. M., Eriksson, M., An, J., Hwang, L. D., Easton, D. F., et al. (2020). Evaluating the role of alcohol consumption in breast and ovarian cancer susceptibility using population-based cohort studies and two-sample Mendelian randomization analyses. *Int. J. Cancer* 148, 1338–1350. doi: 10.1002/ijc.33308

Ooi, B. N. S., Loh, H., Ho, P. J., Milne, R. L., Giles, G., Gao, C., et al. (2019). The genetic interplay between body mass index, breast size and breast cancer risk: a Mendelian randomization analysis. *Int. J. Epidemiol.* 48, 781–794. doi: 10.1093/ije/dyz124

Papadimitriou, N., Dimou, N., Gill, D., Tzoulaki, I., Murphy, N., Riboli, E., et al. (2021). Genetically predicted circulating concentrations of micronutrients and risk of breast cancer: a Mendelian randomization study. *Int. J. Cancer* 148, 646–653. doi: 10.1002/ijc.33246

Qian, F., Wang, S., Mitchell, J., McGuffog, L., Barrowdale, D., Leslie, G., et al. (2019). Height and body mass index as modifiers of breast cancer risk in BRCA1/2 mutation carriers: a Mendelian randomization study. *J. Natl. Cancer Inst.* 111, 350–364. doi: 10.1093/jnci/djy132

Randall, J. C., Winkler, T. W., Kutalik, Z., Berndt, S. I., Jackson, A. U., Monda, K. L., et al. (2013). Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLoS Genet.* 9:e1003500. doi: 10.1371/journal.pgen.1003500

Richardson, T. G., Sanderson, E., Elsworth, B., Tilling, K., and Davey Smith, G. (2020). Use of genetic variation to separate the effects of early and later life adiposity on disease risk: mendelian randomisation study. *BMJ* 369:m1203. doi: 10.1136/bmj.m1203

Ruth, K. S., Day, F. R., Tyrrell, J., Thompson, D. J., Wood, A. R., Mahajan, A., et al. (2020). Using human genetics to understand the disease impacts of testosterone in men and women. *Nat. Med.* 26, 252–258. doi: 10.1038/s41591-020-0751-5

Schumacher, F. R., Al Olama, A. A., Berndt, S. I., Benlloch, S., Ahmed, M., Saunders, E. J., et al. (2018). Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat. Genet.* 50, 928–936. doi: 10.1038/s41588-018-0142-8

Sheehan, N. A., Didelez, V., Burton, P. R., and Tobin, M. D. (2008). Mendelian randomisation and causal inference in observational epidemiology. *PLoS Med.* 5:e177. doi: 10.1371/journal.pmed.0050177

Shu, X., Wu, L., Khankari, N. K., Shu, X. O., Wang, T. J., Michailidou, K., et al. (2018). Associations of obesity and circulating insulin and glucose with breast cancer risk: a Mendelian randomization analysis. *Int. J. Epidemiol.* 48, 795–806. doi: 10.1093/ije/dyy201

Shungin, D., Winkler, T. W., Croteau-Chonka, D. C., Ferreira, T., Locke, A. E., Magi, R., et al. (2015). New genetic loci link adipose and insulin biology to body fat distribution. *Nature* 518, 187–196. doi: 10.1038/nature14132

Tan, V. Y., Yarmolinsky, J., Lawlor, D. A., and Timpson, N. J. (2019). Letter regarding article, "Associations of obesity and circulating insulin and glucose with breast cancer risk: a Mendelian randomization analysis'. *Int. J. Epidemiol.* 48, 1014–1015. doi: 10.1093/ije/dyz013

The 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. *Nature* 526, 68–74. doi: 10.1038/nature15393

Verbanck, M., Chen, C. Y., Neale, B., and Do, R. (2018). Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* 50, 693–698. doi: 10.1038/s41588-018-0099-7

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., et al. (2017). 10 Years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.* 101, 5–22. doi: 10.1016/j.ajhg.2017.06.005

Wang, Y., O'Neil, A., Jiao, Y., Wang, L., Huang, J., Lan, Y., et al. (2019). Sex differences in the association between diabetes and risk of cardiovascular disease, cancer, and all-cause and cause-specific mortality: a systematic review and meta-analysis of 5,162,654 participants. *BMC Med.* 17:136. doi: 10.1186/s12916-019-1355-0

Watts, E. L., Fensom, G. K., Smith Byrne, K., Perez-Cornago, A., Allen, N. E., Knuppel, A., et al. (2020). Circulating insulin-like growth factor-I, total and free testosterone concentrations and prostate cancer risk in 200000 men in UK Biobank. *Int. J. Cancer* 148, 2274–2288. doi: 10.1002/ijc.33416

Yavorska, O. O., and Burgess, S. (2017). MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int. J. Epidemiol.* 46, 1734–1739. doi: 10.1093/ije/dyx034

Yu, X., Wang, T., Chen, Y., Shen, Z., Gao, Y., Xiao, L., et al. (2020). Alcohol drinking and amyotrophic lateral sclerosis: an instrumental variable causal inference. *Ann. Neurol.* 88, 195–198. doi: 10.1002/ana.25721

Yuan, Z., Zhu, H., Zeng, P., Yang, S., Sun, S., Yang, C., et al. (2020). Testing and controlling for horizontal pleiotropy with probabilistic Mendelian randomization in transcriptome-wide association studies. *Nat. Commun.* 11:3861. doi: 10.1038/s41467-020-17668-6

Zeng, P., Wang, T., Zheng, J., and Zhou, X. (2019). Causal association of type 2 diabetes with amyotrophic lateral sclerosis: new evidence from Mendelian randomization using GWAS summary statistics. *BMC Med.* 17:225. doi: 10.1186/s12916-019-1448-9

Zeng, P., Zhao, Y., Qian, C., Zhang, L., Zhang, R., Gou, J., et al. (2015). Statistical analysis for genome-wide association study. *J. Biomed. Res.* 29, 285–297. doi: 10.7555/jbr.29.20140007

Zeng, P., and Zhou, X. (2019a). Causal association between birth weight and adult diseases: evidence from a Mendelian randomization analysis. *Front. Genet.* 10:618. doi: 10.3389/fgene.2019.00618

Zeng, P., and Zhou, X. (2019b). Causal effects of blood lipids on amyotrophic lateral sclerosis: a Mendelian randomization study. *Hum. Mol. Genet.* 28, 688–697. doi: 10.1093/hmg/ddy384