

Title: Selection bias as an explanation for the observed protective association of childhood adiposity with breast cancer

Authors: C M Schooling^{1,2}, K Fei¹, J V Zhao²

¹CUNY School of Public Health, 55 West 125th St, New York, NY 10027, USA

²School of Public Health, The University of Hong Kong, 7 Sassoon Road, Pokfulam, Hong Kong, China

Corresponding author: CM Schooling: cms1@hku.hk, mary.schooling@sph.cuny.edu

Abstract: 200

Word count: 2448

Figures: 1

References: 38

Abstract

Objective: Recalled childhood adiposity is inversely associated with breast cancer observationally, including in Mendelian randomization (MR) studies, questioning its role. Breast cancer studies recruited in adulthood only include survivors of childhood adiposity and breast cancer. We assessed recalled childhood adiposity on participant reported sibling and maternal breast cancer to ensure ascertainment of non-survivors using MR.

Study Design and Setting: We obtained independent strong genetic predictors of recalled childhood adiposity for women and their associations with participant reported own, sibling and maternal breast cancer from UK Biobank genome wide association studies (GWAS). We obtained MR inverse variance weighting estimates.

Results: Childhood adiposity in women was inversely associated with own breast cancer (odds ratio (OR) 0.66, 95% confidence interval (CI) 0.52 to 0.84) but unrelated to participant reported sibling (OR 0.85, 95% CI 0.60 to 1.20) or maternal breast cancer (OR 0.84, 95% CI 0.67 to 1.05 respectively).

Conclusion: Weaker inverse associations of recalled childhood adiposity with breast cancer with more comprehensive ascertainment of cases before recruitment suggests the inverse association of recalled childhood adiposity with breast cancer is due to selection bias arising from preferential selection of survivors. Greater consideration of left truncation in public health relevant causal inferences is warranted.

Keywords: selection bias, survival bias, adiposity, breast cancer, left truncation

Highlights

Recalled childhood adiposity is inversely associated with breast cancer.

Studies of childhood exposures recruited in adulthood are open to left truncation.

Participant reports about family members include deaths before recruitment.

We tested childhood adiposity on sibling breast cancer using Mendelian randomization.

Childhood adiposity with sibling breast cancer was null, suggesting left truncation.

Introduction

Over many years, observational studies have reported recalled childhood adiposity inversely associated with breast cancer,¹⁻⁴ raising questions about the role of adiposity in childhood for girls. Observational studies are open to confounding. As such, these somewhat counter-intuitive findings could be the result of uncontrolled confounding, given the difficulty of enumerating and measuring all potential confounders comprehensively. Recently, a Mendelian randomization (MR) study, which is less open to bias from confounding because it uses genetic proxies of exposure,⁵ also found recalled childhood adiposity inversely associated with breast cancer.⁶ Several plausible reasons have been suggested for childhood adiposity protecting against breast cancer. Greater childhood adiposity could slow pubertal growth and sexual maturation, which could be protective.⁴ Childhood adiposity could also result in a hormonal milieu protective against breast cancer.⁴ Alternatively, the effect of exposure to adiposity during childhood may not be relevant into adulthood,⁶ or may be different, perhaps because the effects of early life experiences are only fully actuated after puberty. A very comprehensive MR study, also using recalled childhood adiposity (at ~57 years), recently reported no biological explanation for the observed inverse association of childhood adiposity with breast cancer,⁷ leaving the inverse association of childhood adiposity with breast cancer an unresolved paradox. Here, we considered another possible explanation: that the observed inverse association of childhood adiposity with breast cancer is bias rather than a target of intervention.

Observational studies, apart from being open to confounding, are also open to selection bias,⁸ particularly for exposures, such as childhood adiposity, that may occur well before recruitment.⁹ Specifically, a study of the effect of childhood adiposity on breast cancer recruited in adulthood is missing the women who died from childhood adiposity before recruitment and the women who died of breast cancer before recruitment, particularly given breast cancer causes death relatively early in adulthood.¹⁰ Only studying those who have survived to recruitment, possibly years after exposure, creates a specific form of selection bias, i.e. survival bias, or depletion of susceptibles,¹¹ because of incomplete ascertainment particularly of exposed cases. Currently, few methods of addressing survival bias exist, although sensitivity analysis is possible.^{12, 13}

Here, we used an MR study to assess the association of childhood adiposity with participant reported sibling and maternal breast cancer to obtain more comprehensive ascertainment of deaths from breast cancer before recruitment, given an MR study reduces bias from confounding.⁵ Using a design based on participant reports of sibling or maternal breast cancer reduces survival bias by including breast cancer cases even if the case died before recruitment of the participant (Figure 1b compared with Figure 1a). So, associations of childhood adiposity

with participant reported sibling or maternal breast cancer should give a less biased estimate for childhood adiposity on breast cancer. We also assessed the potential relevance of survival bias due to death before recruitment from childhood adiposity or breast cancer, from the associations of childhood adiposity and liability to breast cancer with lifespan. The weaker the association for liability to breast cancer the more likely the GWAS is missing deaths before recruitment given breast cancer is a major cause of premature mortality in women.¹⁴

Materials and Methods

MR as a form of instrumental variable analysis requires the assumptions of relevance, independence, and exclusion restriction. To address relevance, we only selected as instruments genome wide significant ($p\text{-value} < 5 \times 10^{-8}$) independent ($r^2 < 0.001$) genetic variants for each exposure. Use of genetic instruments helps ensure independence from confounders.¹⁵ The exclusion restriction assumption is more difficult to satisfy because it can be violated by genetic pleiotropy or selection bias, including survival bias.¹³ We assessed genetic pleiotropy by sensitivity analysis. We addressed potential selection bias for childhood adiposity on breast cancer by using participant reported sibling and maternal breast cancer as outcomes, and from the relevance of potentially biasing pathways (Figure 1), i.e., childhood adiposity and breast cancer on survival to recruitment proxied by lifespan. We used two-sample MR, where possible. Some questions could only be addressed using one-sample MR which can be biased even for very large samples.¹⁶ So, we assessed the validity of one-sample MR estimates from sample size and relevant diagnostics, such as I^2_{GX} .¹⁶

Study design

This is mainly a one-sample MR study using genome wide association studies (GWAS) of the UK Biobank. The UK Biobank is a population-based cohort of half a million people, intended to be recruited at age 40 to 69 years, average age ~57 years, from 22 sites in Great Britain in 2006 to 2010.¹⁷ The participants completed extensive questionnaires, as well as undergoing tests and providing samples for genotyping and other investigations. The UK Biobank includes information on childhood adiposity (relative body size at 10 years), breast cancer in self and family members (siblings and mothers), as well as sex-specific parental survival (age at death or current age). Quality controlled GWAS for up to ~13.5 genetic variants are publicly available, both sex-combined¹⁸ and sex-specific¹⁹ (<http://www.nealelab.is/uk-biobank>). We also used the largest publicly available GWAS of self breast cancer from the Breast Cancer Association Consortium (BCAC),²⁰ as used previously,²¹ which is largely based on case-control studies from Europe and North America.

Data sources

Childhood adiposity

We obtained genetic instruments for childhood adiposity from a UK Biobank GWAS of self-reported comparative childhood body size at age 10 years (thinner, about average, plumper) from the UK Biobank,¹⁷ concerning 191344 white British women adjusted for principal components, age and age² (<http://www.nealelab.is/uk-biobank>).

Breast cancer

We used a quality controlled GWAS of participant reported sibling breast cancer (cases=16586, non-cases=345233) and mothers breast cancer (cases=35102, non-cases=388356) of white Europeans from the UK Biobank,¹⁸ obtained using linear regression after adjusting for principal components. These estimates were converted from probability into log odds using an approximation.²² Given siblings and mother only share half their genetic endowment with the participant, these estimates were also doubled. For comparison, we also used a similarly obtained GWAS of self breast cancer from the UK Biobank (cases=10303, non-cases=452630) and the BCAC GWAS of self breast cancer (cases=122977, controls=105974) largely in Europeans which was quality controlled and adjusted for country and principal components.²⁰

Lifespan

We used a GWAS of mother's attained age from the UK Biobank for lifespan,¹⁹ based on reports of mother's attained age (age at death or current age), which is partly heritable.²³ Only white Europeans were included.²⁴ Deaths of mothers before 57 years and adopted participants were excluded.²⁴ Quality controlled genetic associations with Martingale residuals from Cox proportional hazards were adjusted for age, assessment centre and array type.²⁴ Estimates were converted into life years using an established approximation.²⁵

Statistical analysis

Genetic instruments

We selected genome wide significant (p -value $< 5 \times 10^{-8}$) independent ($r^2 < 0.001$) genetic variants as instruments for all exposures considered. As a measure of instrument strength, we obtained each F-statistic as the estimate for genetic variant on exposure divided by its variance.²⁶ We aligned genetic variants across studies on the same allele letter, using effect allele frequency as necessary after dropping palindromic variants with minor allele frequency > 0.42 .

Analysis

We used inverse variance weighting (IVW) as the main analysis. We used the weighted median (WM) and MR-Egger (MRE) as sensitivity analysis to assess potential violation of the exclusion restriction assumption by pleiotropy or selection bias. The WM gives a valid estimate as long as at least 50% of the weight is based on valid instruments.²⁷ MRE gives valid estimate when the genetic instruments do not affect confounders of exposure on outcome,²⁶ but has low power.

MR-Base was used to extract instruments and align genetic variants across studies,¹⁸ the MendelianRandomization R package was used to obtain estimates. All analyses were conducted using R (version 4.1.2).²⁸ This study only uses genetic summary statistics created from information and materials previously collected with informed consent.

Results

At least 75 genetic variants with F-statistics greater than 10 and average F-statistic >66 were obtained for childhood adiposity in women and were available for all outcomes (Table 1). Childhood adiposity was not clearly related to UK Biobank participant reported sibling breast cancer or to UK Biobank participant reported mother's breast cancer. In contrast, childhood adiposity was strongly inversely associated with self breast cancer from UK Biobank and BCAC (Table 1).²⁰ Sensitivity analysis, i.e., the MR Egger intercept, indicated that the estimates for childhood adiposity on breast cancer from BCAC might be invalid (Table 3) due to pleiotropy or selection bias.

As expected, childhood adiposity in women was strongly inversely associated with lifespan in women (Table 2). MR estimates were similar from all sensitivity methods, and the MR-Egger intercept p-value did not indicate pleiotropy (Table 2).

Associations of liability to breast cancer with lifespan differed by breast cancer study (Table 3), with the greatest magnitude of association for participant reported sibling breast cancer from the UK Biobank and the smallest magnitude of association for BCAC (Table 3). The F-statistics were adequate. Sensitivity analysis gave a similar interpretation, with no evidence of pleiotropy from the MRE intercepts (Table 3).

Discussion

Childhood body size was not clearly associated with breast cancer using participant report of siblings' or mothers' breast cancer to obtain more complete ascertainment of lifetime breast cancer. Moreover, we showed the existence of biasing pathways when considering effects of a

harmful early life exposure on a disease that can cause death relatively early in adulthood. Collectively, this study shows that the inverse association of childhood adiposity with breast cancer could be an artefact of selection bias arising from only selecting survivors. Moreover, the finding of no association of childhood adiposity with participant reported sibling's breast cancer or mother's breast cancer is consistent with a recent study which found no physiological reason why childhood adiposity should protect against breast cancer,⁷ but, parsimoniously, suggests the explanation is a bias.

Previous studies have emphasized the importance, when assessing potentially causal effects, of ensuring the exposure starts at or after recruitment to avoid biased estimates,⁹ as well as the importance of counting exposure time correctly for the exposed and unexposed.²⁹ The issues raised by some potential participants being missing from a study because of prior death due to the exposure or outcome, or a competing risk of the outcome³⁰ is known,³¹⁻³³ sometimes as left truncation, and is often thought to have most relevance for studies in the elderly.³¹⁻³³ In contrast, population representative studies are widely considered to be a keystone of scientific research.³⁴ So bias in associations for long-term harmful exposures in studies only representing survivors of the original birth cohorts who formed the population may seem counter-intuitive.

Despite using MR to reduce confounding, using participant reported sibling and maternal breast cancer to address selection bias and conducting additional analysis to check the conditions for selection bias, this study has limitations. First, childhood adiposity was self-reported in adulthood in three groups, so may not be entirely accurate, possibly reducing precision which is compensated for by the large sample size. Second, lifespan was based on participant (adult child) reported maternal attained age which could be open to measurement error, potentially biasing towards the null. However, reports about others can be more reliable than self-reports³⁵ and mother's age is also important. Mothers who died early (before age 57 years) were excluded, likely biasing towards the null or the reverse of the true association. Third, sibling's and mother's breast cancer could be reported imprecisely, which would bias estimates towards the null. Mother's breast cancer could also be open to survival bias due to competing risk, where in older women death from cardiovascular disease might preclude the occurrence of breast cancer. Nevertheless, liability to sibling breast cancer had the largest association with life years lost, as expected. Fourth, publicly available GWAS of sibling breast cancer by sex (of the sibling) are not available, making the magnitude of the estimates for sibling breast cancer subject to several assumptions. Fifth, the UK Biobank is not population representative, which would be a source of bias for a descriptive study but is less likely to be a source of bias for MR studies unless the participants were selected on genetic endowment and the outcomes. The BCAC

GWAS of breast cancer is a meta-analysis of case series, case-control studies and cohort studies carefully designed to avoid genetic confounding, but less explicitly designed to address selection bias.²⁰ Sixth, MR studies rest on the assumptions of relevance, independence and exclusion. The genetic instruments strongly predicted each exposure, with high F-statistics. Random allocation of genetic material at conception reduces the risk of the genetic instruments being associated with confounders. However, there remains a possibility of violating the exclusion restriction assumption by selecting on childhood BMI. Such bias is not thought to be very influential,³⁶ but would bias towards a protective effect of childhood BMI. Seventh, one-sample MR, which can be biased towards the confounded estimate, was largely used. However, differences were found between estimates for childhood adiposity on breast cancer from participant self-reports and participant reports about their family (siblings and mother) in the UK Biobank when confounding is likely to be similar for self, siblings and mother. Moreover, the large sample sizes suggest any such bias is likely to be minimal for IVW and WM estimates.¹⁶

Here, we demonstrate that when using observational studies to examine the effect of exposures that start before recruitment it is important to consider whether selection bias due to differential survival might have occurred because of inadvertent exclusion of deaths before recruitment. Use of MR studies with suitable proxy outcomes,³⁷ such as siblings or parents, may help identify bias arising from inadvertently focusing on survivors. However, in this context it is vital to obtain information about all siblings/mothers regardless of mortality status rather than only including or recruiting those who survived to recruitment, which would not address selection bias due to only selecting survivors. Overall, our study draws attention to the importance of considering whether studies recruited in adulthood, particularly of lifetime exposures, such as genetics, could be biased by almost inevitable selection on survival. Bias is likely to be most marked when selecting on survival of exposure and outcome,³⁸ or on survival of exposure and a competing risk of the outcome.^{13, 30} Survival bias is particularly relevant for studies, such as genetic studies, where the exposure often commences well before recruitment and the outcome is a disease state, where missing cases may considerably distort the comparisons. In contrast, for continuous outcomes missing a few people at the extremes of the distribution is unlikely to affect comparisons of central tendency.

Conclusion

Observations of childhood adiposity associated with a lower risk of breast cancer could be an artefact of selection bias. Assessing the risk of selection bias, due to inadvertently only including survivors, by using controls, such as participant reports on family members, and sensitivity

analysis could ensure resources are focused on valid questions, and doubt is not cast on important targets of intervention, such as childhood adiposity.

Ethics Approval

This study only uses publicly available summary statistics, so does not require ethical approval.

Author contributions:

CMS designed the study in consultation with KF. CMS conducted the initial analysis. JVZ and KF reviewed and checked the analysis and results. CMS wrote the initial draft, which JVZ and KF reviewed thoroughly for important intellectual content. All authors gave final approval to this version and are accountable for all aspects of the work.

Data Statement:

The data used here is available in MR-Base, apart from the GWAS of childhood adiposity in women which is available from Neale Lab <http://www.nealelab.is/uk-biobank>

Funding: None

Acknowledgements: None

Conflict of Interest/competing interests: None declared

Author Contributions: CMS conceptualized the study and the methodology, conducted the initial analysis and wrote the original draft. All authors discussed and reviewed the study methodology in detail. KF and JVZ checked the analysis, edited and reviewed the final version critically for important intellectual content and approved the final version for submission.

Replication can be achieved using MR Base <https://www.mrbase.org>, publicly available data and R code available on request

Table 1: MR estimates in women for childhood adiposity from the UK Biobank on breast cancer from UK Biobank participant reports about siblings and mothers and on self breast cancer from the UK Biobank and BCAC

Units	Exposure Description	Instruments # (mean f-statistic)	Breast cancer outcome for Subject	Study used	Method	Odds ratio	95% CI	p-value	MR-Egger intercept p, Q(p), I ² _{GX}
Childhood adiposity per difference between groups	Recalled comparative body size at 10 years (thinner, about average, plumper) http://www.nealelab.is/uk-biobank	79 (67.1)	Sibling ¹	UK Biobank (ukb-b-12227) ²	IVW	0.85	0.60 to 1.20	0.36	0.10, 102.0 (0.03), 90.4%
					WM	1.02	0.62 to 1.70	0.93	
					MRE	1.52	0.70 to 3.31	0.29	
		79 (67.1)	Mother	UK Biobank (ukb-b-13584)	IVW	0.84	0.67 to 1.05	0.13	0.12, 106.9 (0.01), 90.6%
					WM	0.67	0.50 to 0.92	0.01	
					MRE	0.58	0.35 to 0.97	0.04	
		77 (67.4)	Self	UK Biobank (ukb-b-16890) ¹	IVW	0.66	0.52 to 0.84	0.001	0.58, 140.7(0), 90.4%
					WM	0.67	0.50 to 0.90	0.01	
					MRE	0.58	0.33 to 0.99	0.046	
		75 (66.5)	Self	BCAC (ieu-a-1126)	IVW	0.66	0.57 to 0.77	4.9e-8	0.001, 226.4(0), 90.2%
					WM	0.60	0.52 to 0.70	8.1-10	
					MRE	0.41	0.29 to 0.56	3.5e-8	

IVW = Inverse variance weight; WM = Weighted median; MRE = MR-Egger.

¹ Given the denominator relates to male and female siblings when converting to an odds ratio we halved the denominator and also doubled the estimate, i.e., we assumed half the participants had a female sibling given the family size was about two.

Table 2: MR estimates in women for childhood adiposity on lifespan (life years) in the UK Biobank

Units	Subject	Exposure		n	Method	Instrument # (mean f-statistic)	Outcome – Lifespan (ebi-a-GCST006696)			MR-Egger intercept p-value, Q(p), I ² _{GX}
		Description					Life years ¹	95% CI	p-value	
Childhood adiposity per difference between groups ²	Self	Comparative body size at 10 years		191344	IVW	79 (67.1)	-1.67	-2.71 to -0.64	0.002	0.22, 174 (0), 90.3%
		http://www.nealelab.is/uk-biobank			WM		-1.25	-2.44 to -0.06	0.04	
					MRE		-2.98	-5.31 to -0.65	0.01	

IVW = Inverse variance weight; WM = Weighted median; MRE = MR-Egger.

¹ Years of life obtained from the log protection ratio multiplied by 2.5863 to account for children only sharing half their genetic endowment with their mothers and then by 10 as a verified actuarial rule of thumb ¹⁹

² Thinner, about average, plumper compared to others at age 10 years

Table 3: MR estimates in women for liability to breast cancer from UK Biobank participant reports about siblings and mothers and for self from the UK Biobank and BCAC on lifespan (life years)

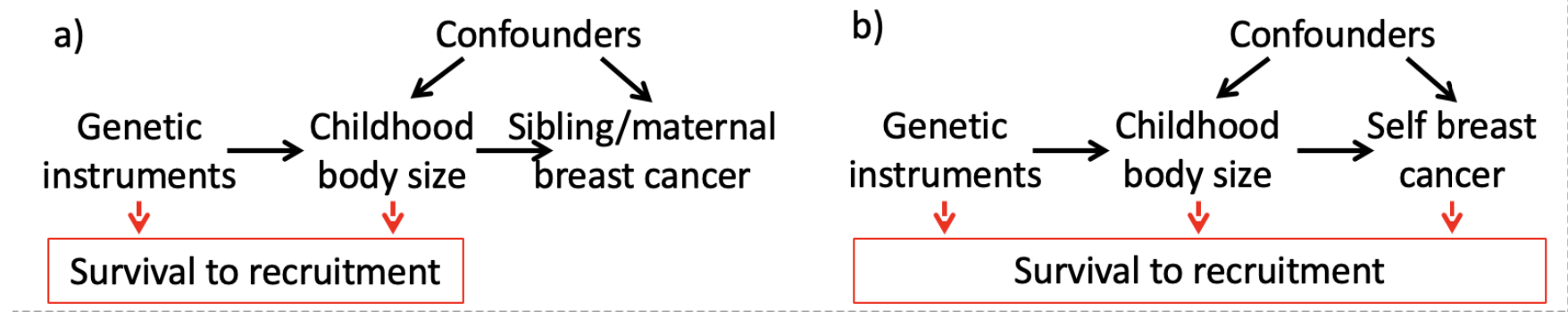
Units	Subject	Exposure Description	Cases/participants	Instrument #s (mean f-statistic)	Method	Outcome – Longevity (ebi-a-GCST006696) Life years ¹	95% CI	p-value	MR-Egger intercept p-value, Q(p), I ² _{Gx}
Liability to breast cancer in log odds	Sibling	UK Biobank (ukb-b-12227)	16586/ 361809 ²	6 (49.2)	IVW	-0.80	-1.13 to -0.47	1.8E-06	0.17, 4.8 (0.31), 0%
					WM	-0.76	-1.20 to -0.32	0.001	
					MRE	0.10	-1.21 to 1.41	0.89	
	Mother	UK Biobank (ukb-b-13584)	35102/ 423458	12 (70.7)	IVW	-0.55	-0.89 to -0.20	0.002	0.66, 21.2 (0.02), 86.8%
					WM	-0.69	-1.04 to -0.34	0.0001	
					MRE	-0.75	-1.76 to -0.25	0.14	
	Self	UK Biobank (ukb-b-16890)	10303/ 462933	17 (78.0)	IVW	-0.57	-0.89 to -0.26	0.0004	0.46, 32.0 (0.01), 90.1%
					WM	-0.56	-0.90 to -0.22	0.001	
					MRE	-0.85	-1.64 to -0.05	0.04	
	Self	BCAC (ieu-a-1126)	122977/ 228951	122 (87.1)	IVW	-0.53	-0.74 to -0.31	2.8E-06	0.70, 193(0), 93.1%
					WM	-0.43	-0.72 to -0.13	0.005	
					MRE	-0.61	-1.10 to -0.12	0.014	

IVW = Inverse variance weight; WM = Weighted median; MRE = MR-Egger.

¹ Years of life obtained from the log protection ratio multiplied by 2.5863 to account for children only sharing half their genetic endowment with their mothers and then by 10 as a verified actuarial rule of thumb¹⁹

² Given the denominator relates to male and female siblings when converting to an odds ratio we halved the denominator and also doubled the estimate, i.e., we assumed half the participants had a female sibling given the family size was about two.

Figure 1: Directed acyclic graph showing that a Mendelian randomization study of childhood adiposity on self breast cancer recruited in middle to older age is biased by selection on genetic endowment and surviving breast cancer (Figure 1b) but a study of participant reported sibling or maternal breast cancer is likely less biased because it does not select on surviving breast cancer to recruitment (Figure 1a). A box indicates selection on an attribute.



References

1. Yang TO, Cairns BJ, Pirie K, Green J, Beral V, Floud S, et al. Body size in early life and the risk of postmenopausal breast cancer. *BMC Cancer*. 2022;22(1):232.
2. Llewellyn A, Simmonds M, Owen CG, Woolacott N. Childhood obesity as a predictor of morbidity in adulthood: a systematic review and meta-analysis. *Obes Rev*. 2016;17(1):56-67.
3. Byun D, Hong S, Ryu S, Nam Y, Jang H, Cho Y, et al. Early-life body mass index and risks of breast, endometrial, and ovarian cancers: a dose-response meta-analysis of prospective studies. *Br J Cancer*. 2022;126(4):664-72.
4. Baer HJ, Colditz GA, Rosner B, Michels KB, Rich-Edwards JW, Hunter DJ, et al. Body fatness during childhood and adolescence and incidence of breast cancer in premenopausal women: a prospective cohort study. *Breast Cancer Res*. 2005;7(3):R314-25.
5. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol*. 2003;32(1):1-22.
6. Richardson TG, Sanderson E, Elsworth B, Tilling K, Davey Smith G. Use of genetic variation to separate the effects of early and later life adiposity on disease risk: mendelian randomisation study. *Bmj*. 2020;369:m1203.
7. Vabistsevits M, Davey Smith G, Sanderson E, Richardson TG, Lloyd-Lewis B, Richmond RC. Deciphering how early life adiposity influences breast cancer risk using Mendelian randomization. *Commun Biol*. 2022;5(1):337.
8. Bareinboim E, Pearl J. Causal inference and the data-fusion problem. *Proc Natl Acad Sci U S A*. 2016;113(27):7345-52.
9. Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *J Clin Epidemiol*. 2016;79:70-5.
10. Kesteloot H, Decramer M. Age at death from different diseases: the flemish experience during the period 2000-2004. *Acta Clin Belg*. 2008;63(4):256-61.
11. Stovitz SD, Banack HR, Kaufman JS. 'Depletion of the susceptibles' taught through a story, a table and basic arithmetic. *BMJ Evid Based Med*. 2018;23(5):199.
12. Smith LH, VanderWeele TJ. Bounding Bias Due to Selection. *Epidemiology*. 2019;30(4):509-16.
13. Schooling CM, Lopez PM, Yang Z, Zhao JV, Au Yeung SL, Huang JV. Use of Multivariable Mendelian Randomization to Address Biases Due to Competing Risk Before Recruitment. *Front Genet*. 2020;11:610852.
14. Fitzmaurice C, Abate D, Abbasi N, Abbastabar H, Abd-Allah F, Abdel-Rahman O, et al. Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-Years for 29 Cancer Groups, 1990 to 2017: A Systematic Analysis for the Global Burden of Disease Study. *JAMA Oncol*. 2019;5(12):1749-68.
15. Smith GD, Lawlor DA, Harbord R, Timpson N, Day I, Ebrahim S. Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology. *PLoS Med*. 2007;4(12):e352.
16. Minelli C, Del Greco MF, van der Plaats DA, Bowden J, Sheehan NA, Thompson J. The use of two-sample methods for Mendelian randomization analyses on single large datasets. *Int J Epidemiol*. 2021;50(5):1651-9.
17. Collins R. What makes UK Biobank special? *Lancet*. 2012;379(9822):1173-4.

18. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife*. 2018;7.
19. Pilling LC, Kuo CL, Sicinski K, Tamosauskaite J, Kuchel GA, Harries LW, et al. Human longevity: 25 genetic loci associated in 389,166 UK biobank participants. *Aging (Albany NY)*. 2017;9(12):2504-20.
20. Michailidou K, Lindström S, Dennis J, Beesley J, Hui S, Kar S, et al. Association analysis identifies 65 new breast cancer risk loci. *Nature*. 2017;551(7678):92-4.
21. . !!! INVALID CITATION !!! 6, 7.
22. Lloyd-Jones LR, Robinson MR, Yang J, Visscher PM. Transformation of Summary Statistics from Linear Mixed Model Association on All-or-None Traits to Odds Ratio. *Genetics*. 2018;208(4):1397-408.
23. Gudmundsson H, Gudbjartsson DF, Frigge M, Gulcher JR, Stefánsson K. Inheritance of human longevity in Iceland. *Eur J Hum Genet*. 2000;8(10):743-9.
24. Pilling LC, Tamosauskaite J, Jones G, Wood AR, Jones L, Kuo CL, et al. Common conditions associated with hereditary haemochromatosis genetic variants: cohort study in UK Biobank. *Bmj*. 2019;364:k5222.
25. Timmers PR, Mounier N, Lall K, Fischer K, Ning Z, Feng X, et al. Genomics of 1 million parent lifespans implicates novel pathways and common diseases and distinguishes survival chances. *Elife*. 2019;8.
26. Bowden J, Del Greco MF, Minelli C, Davey Smith G, Sheehan NA, Thompson JR. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I² statistic. *International Journal of Epidemiology*. 2016;45(6):1961-74.
27. 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(4):304-14.
28. R Core Team. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2020.
29. Suissa S. Immortal time bias in pharmaco-epidemiology. *Am J Epidemiol*. 2008;167(4):492-9.
30. Thompson CA, Zhang ZF, Arah OA. Competing risk bias to explain the inverse relationship between smoking and malignant melanoma. *Eur J Epidemiol*. 2013;28(7):557-67.
31. Hernán MA, Alonso A, Logroscino G. Cigarette smoking and dementia: potential selection bias in the elderly. *Epidemiology*. 2008;19(3):448-50.
32. Vansteelandt S, Dukes O, Martinussen T. Survivor bias in Mendelian randomization analysis. *Biostatistics*. 2018;19(4):426-43.
33. Vansteelandt S, Walter S, Tchetgen Tchetgen E. Eliminating Survivor Bias in Two-stage Instrumental Variable Estimators. *Epidemiology*. 2018;29(4):536-41.
34. Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. *Soc Sci Med*. 2018;210:2-21.
35. Mathiowetz NA, Groves RM. The effects of respondent rules on health survey reports. *Am J Public Health*. 1985;75(6):639-44.
36. Gkatzionis A, Burgess S. Contextualizing selection bias in Mendelian randomization: how bad is it likely to be? *Int J Epidemiol*. 2019;48(3):691-701.

37. Davey Smith G, Leary S, Ness A, Lawlor DA. Challenges and novel approaches in the epidemiological study of early life influences on later disease. *Adv Exp Med Biol.* 2009;646:1-14.
38. Westreich D. Berkson's bias, selection bias, and missing data. *Epidemiology.* 2012;23(1):159-64.