University Library

MINERVA ACCESS

A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Li, S;Dite, GS;MacInnis, RJ;Bui, M;Nguyen, TL;Esser, VFC;Ye, Z;Dowty, JG;Makalic, E;Sung, J;Giles, GG;Southey, MC;Hopper, JL

Title:
Causation and familial confounding as explanations for the associations of polygenic risk scores with breast cancer: Evidence from innovative ICE FALCON and ICE CRISTAL analyses

Date:
2024-12-01

Citation:
Li, S., Dite, G. S., MacInnis, R. J., Bui, M., Nguyen, T. L., Esser, V. F. C., Ye, Z., Dowty, J. G., Makalic, E., Sung, J., Giles, G. G., Southey, M. C. & Hopper, J. L. (2024). Causation and familial confounding as explanations for the associations of polygenic risk scores with breast cancer: Evidence from innovative ICE FALCON and ICE CRISTAL analyses. Genetic Epidemiology, 48 (8), pp.401-413. https://doi.org/10.1002/gepi.22556.

Persistent Link:
https://hdl.handle.net/11343/345434

License:
CC BY

Check for updates

Received: 31 July 2023 | Accepted: 23 February 2024

DOI: 10.1002/gepi.22556

RESEARCH ARTICLE

Genetic Epidemiology

OFFICIAL JOURNAL INTERNATIONAL GENETIC EPIDEMIOLOGY SOCIETY www.geneticepi.org

WILEY

Causation and familial confounding as explanations for the associations of polygenic risk scores with breast cancer: Evidence from innovative ICE FALCON and ICE CRISTAL analyses

Shuai Li[1,2,3,4] | **Gillian S. Dite**[1,5] | **Robert J. MacInnis**[1,6] | **Minh Bui**[1] | **Tuong L. Nguyen**[1] | **Vivienne F. C. Esser**[1] | **Zhoufeng Ye**[1] | **James G. Dowty**[1] | **Enes Makalic**[1] | **Joohon Sung**[7,8,9] | **Graham G. Giles**[1,3,6] | **Melissa C. Southey**[3,6,10] | **John L. Hopper**[1]

[1]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Carlton, Victoria, Australia

[2]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

[3]Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, Victoria, Australia

[4]Murdoch Children's Research Institute, Royal Children's Hospital, Parkville, Victoria, Australia

[5]Genetic Technologies Ltd., Fitzroy, Victoria, Australia

[6]Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, Victoria, Australia

[7]Division of Genome and Health Big Data, Department of Public Health Sciences, Graduate School of Public Health, Seoul National University, Seoul, Korea

[8]Genomic Medicine Institute, Seoul National University, Euigwahakgwan #402, Seoul National University College of Medicine, 103, Daehak-ro, Jongno-gu, Seoul, South Korea

[9]Institute of Health and Environment, Seoul National University, 1st GwanakRo, Seoul, South Korea

[10]Department of Clinical Pathology, The University of Melbourne, Parkville, Victoria, Australia

Correspondence
John L. Hopper, Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, 207 Bouverie St., Carlton, VIC 3053, Australia.
Email: j.hopper@unimelb.edu.au

Funding information
National Cancer Institute; National Health and Medical Research Council, Grant/Award Number: GNT2024867; New South Wales Cancer Council, the Victorian Health Promotion Foundation; Victorian Breast Cancer Research Consortium, Cancer Australia; National Breast Cancer Foundation; U.S. National Cancer

Abstract

A polygenic risk score (PRS) combines the associations of multiple genetic variants that could be due to direct causal effects, indirect genetic effects, or other sources of familial confounding. We have developed new approaches to assess evidence for and against causation by using family data for pairs of relatives (Inference about Causation from Examination of FAmiliaL CONfounding [ICE FALCON]) or measures of family history (Inference about Causation from Examining Changes in Regression coefficients and Innovative STatistical AnaLyses [ICE CRISTAL]). Inference is made from the changes in regression coefficients of relatives' PRSs or PRS and family history before and after adjusting for each other. We applied these approaches to two breast cancer PRSs and multiple studies and found that (a) for breast cancer

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2024 The Authors. *Genetic Epidemiology* published by Wiley Periodicals LLC.

Institute, Grant/Award Number: UM1 CA164920; Cancer Council Victoria and VicHealth; Cancer Council Victoria; National Research Foundation of Korea (NRF), Grant/Award Numbers: 2020R1A2C2101041, 2023R1A2C2002925

diagnosed at a young age, for example, <50 years, there was no evidence that the PRSs were causal, while (b) for breast cancer diagnosed at later ages, there was consistent evidence for causation explaining increasing amounts of the PRS-disease association. The genetic variants in the PRS might be in linkage disequilibrium with truly causal variants and not causal themselves. These PRSs cause minimal heritability of breast cancer at younger ages. There is also evidence for nongenetic factors shared by first-degree relatives that explain breast cancer familial aggregation. Familial associations are not necessarily due to genes, and genetic associations are not necessarily causal.

KEYWORDS

breast cancer, causation, familial confounding, family data, ICE CRISTAL, ICE FALCON, siblings

1 | INTRODUCTION

A polygenic risk score (PRS) combines the associations of multiple genetic variants with a given trait. Although a PRS—like any other risk factor—can be used to predict the likelihood of having the trait now or in the future, is there a causal aspect to this association that could be used to inform prevention and aetiological research? Use in the literature of phrases like 'confer risk' or 'convey risk' imply a causal interpretation, but is this just the writer who is doing the conferring (labelling) and conveying (putting ideas into the reader's mind) rather than the implicated genes and/or genetic variants themselves being biologically relevant to the disease in question?

Associations can always be due to confounding, whereby there is an external factor (confounder) that influences both the exposure and the outcome. Genetic associations, and therefore genetic predictions, could therefore be due to familial confounding in which the external factor is also correlated to some extent in relatives. For example, a PRS-trait association does not imply that all or even any of the measured variants are causal because variants can be 'tagged' by truly causal variants. That is, on a population basis a truly causal genetic variant could be a confounder for the trait association of a measured genetic variant included in the PRS.

It is also possible for a PRS-trait association to be due at least in part to other sources of familial confounding. Some possibilities include (i) population stratification, in which ancestry is correlated with both phenotypes and genotypes due to geographic or regional differences in allele frequency relating to the trait of interest; (ii) assortative as distinct from random mating, whereby people mate with those of more similar phenotypes for whatever reason resulting in parents being more similar for the genetic aspects underlying those phenotypes than by chance and (iii) indirect genetic effects in which the expression of a parental genotype on the parental phenotype directly affects the offspring phenotype (Brumpton et al., 2020). There could be other reasons as well.

For breast cancer, both a 77-single-nucleotide polymorphism (SNP) PRS (Mavaddat et al., 2015) and more recently a 313-SNP PRS (Mavaddat et al., 2019) have been shown to predict disease risk with the odds ratio per cross-sectional standard deviation of the PRS (odds ratio [OR]) estimated to be 1.49 (95% confidence interval [CI]: 1.44, 1.56) and 1.61 (95% CI: 1.57, 1.65), respectively. The 77-SNP PRS was created using a significance threshold of $p < 5 \times 10^{-8}$ and a sample size of 33,673 cases and 33,381 controls, while the 313-SNP was created using $p < 10^{-5}$ and a sample size 94,075 cases and 75,017 controls. Perhaps the truly causal variants are more likely to be captured by PRSs created from studies that use more stringent thresholds, or apply to different subtypes of disease related to age at diagnosis, hormone-receptor status and so on. To the best of our knowledge, the causal aspects of the breast cancer PRSs have not been addressed.

Mendelian randomisation, which uses genetic variants associated with exposures as presumed instrumental variables to assess causal evidence, cannot be used to investigate the causality of a PRS. This is because the PRS is the exposure underlying this research question and is genetic itself—there are no genetic instrumental variables for a PRS.

Family designs have historically been used to investigate the causes of variation (i.e. familial and nonfamilial associations) in measured nonbinary traits. They address the strengths of familial associations, not necessarily the issue of causes per se.

10982272, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/gepi.22556 by National Health And Medical Research Council, Wiley Online Library on [02/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

We have developed innovative methods to assess evidence for and against causation between an exposure and an outcome that apply regression analyses to family data. The first method is called Inference about Causation from Examination of FAmiliaL CONfounding (ICE FALCON) (Li et al., 2020), which uses data from pairs of relatives, such as twins or siblings. The exposure of the relative is used as a measured proxy instrumental variable for assessing causal evidence but without the assumption that there are no other causal pathways from the relative's exposure to the outcome. Importantly, inference using ICE FALCON does not come from fitting one model, but instead from fitting multiple models and studying the changes in pairs of regression coefficients. ICE FALCON analysis has provided causal evidence for numerous exposures and outcomes (Bui et al., 2013, 2020; Dite et al., 2008; Dongmeng et al., 2022; Hong et al., 2023; Hopper et al., 2012; Li et al., 2018, 2019, 2020; Nissen et al., 2023; Stone et al., 2012; Tian et al., 2023; Wu et al., 2022; Zheng et al., 2022).

A more recent method is the Inference about Causation from Examining Changes in Regression coefficients and Innovative STatistical AnaLyses (ICE CRISTAL) (Li & Hopper, 2023), which in effect extends ICE FALCON by using the relative's outcome (i.e. a person's family history, however defined) as the proxy instrumental variable. ICE CRISTAL is therefore more generally applicable as it does not rely on the exposure of the relative to be measured and can use data for unrelated individuals. ICE CRISTAL previously provided evidence consistent with mammographic density having a causal effect on breast cancer risk (Martin et al., 2010), although this insight was not made at the time. More details about these methods can be found in the Methods section.

In this study, we aimed to investigate the evidence for causality of the breast cancer PRSs using the ICE FALCON and ICE CRISTAL methods. We investigated the causal effects overall, and by age at diagnosis, using data from three cohorts and published data.

2 | METHODS

2.1 | ICE FALCON analysis

Figure 1 shows two possible causal diagrams between PRS and breast cancer for pairs of relatives: (1) PRS has a causal effect on breast cancer (Figure 1a), and (2) the association between PRS and breast cancer is due to familial confounding, that is, there is no causation (Figure 1b). ICE FALCON fits three models:

1. $E\,(\mathrm{BC}_{\mathrm{self}}) = \alpha + \beta_{\mathrm{self}}\mathrm{PRS}_{\mathrm{self}}$, which estimates the marginal association between a woman's breast cancer and her own PRS (β_{self}).

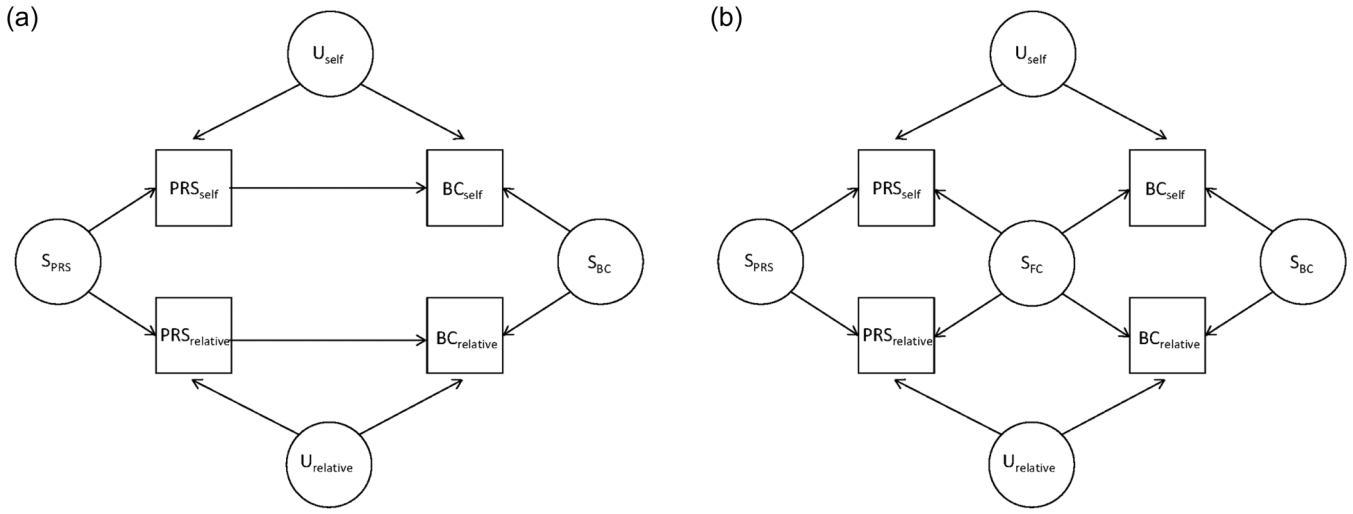


FIGURE 1 Possible causal diagrams between PRS and breast cancer for pairs of relatives: (a) causation and (b) familial confounding. PRS represents a polygenic risk score and BC represents breast cancer. S denotes the combinations of all unmeasured factors (genetic and/or nongenetic) that affect both relatives: S_{PRS} represents those causing the PRS only, S_{BC} represents those causing breast cancer only, and S_{FC} represents those causing both the PRS and breast cancer (i.e., familial confounders). U denotes all the unmeasured individual-specific confounders between PRS and breast cancer that are not shared by relatives. For the purposes of explanation, 'self' refers to a woman and 'relative' refers to the woman's relative but recognise that these labels can be swapped and both relatives within a pair are included in the analyses.

10982272, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ggpi.22556 by National Health And Medical Research Council, Wiley Online Library on [02/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2. $E(\mathrm{BC}_{\mathrm{self}}) = \alpha + \beta_{\mathrm{relative}}\mathrm{PRS}_{\mathrm{relative}}$, which estimates the marginal association between a woman's breast cancer and her relative's PRS ($\beta_{\mathrm{relative}}$).
3. $E(\mathrm{BC}_{\mathrm{self}}) = \alpha + \beta'_{\mathrm{self}}\mathrm{PRS}_{\mathrm{self}} + \beta'_{\mathrm{relative}}\mathrm{PRS}_{\mathrm{relative}}$, which estimates the conditional association between a woman's breast cancer with her own PRS given her relative's PRS (β'_{self}), and the conditional association between a woman's breast cancer with her relative's PRS given her own PRS ($\beta'_{\mathrm{relative}}$).

Observing that $\mathrm{PRS}_{\mathrm{self}}$ association does not change after conditioning on $\mathrm{PRS}_{\mathrm{relative}}$ (i.e. $\beta'_{\mathrm{self}} - \beta_{\mathrm{self}} \sim 0$) but that $\mathrm{PRS}_{\mathrm{relative}}$ association decreases to the null after conditioning on $\mathrm{PRS}_{\mathrm{self}}$ (i.e. $\beta'_{\mathrm{relative}} \sim 0$ and $\beta'_{\mathrm{relative}} - \beta_{\mathrm{relative}} < 0$) is consistent with PRS having a causal effect on breast cancer (Figure 1a). Observing that both $\mathrm{PRS}_{\mathrm{self}}$ association and $\mathrm{PRS}_{\mathrm{relative}}$ association decrease to the same extent after conditioning on each other (i.e. $\beta'_{\mathrm{self}} - \beta_{\mathrm{self}} \sim \beta'_{\mathrm{relative}} - \beta_{\mathrm{relative}}$) is consistent with there being familial confounding (Figure 1b). For a scenario in which there is a mixture of causation and familial confounding, the results are expected to be a mixture of the results of the two specific scenarios. That is, asymmetry in the changes in regression coefficients is indicative of causation, even in a setting where there is also familial confounding. More details about ICE FALCON can be found in Li et al. (2020).

2.2 | Application of ICE FALCON

We applied ICE FALCON to genotyped sister pairs from the UK Biobank. As in Brumpton et al. (2020), genetic data were used to identify full siblings among UK Biobank participants, resulting in 41,496 siblings in 20,136 families. We selected siblings who were female and unaffected by breast cancer at enrolment, and families with at least two sisters, resulting in 13,696 sisters in 6847 families. For families with more than two sisters, we generated all possible sister pairs. As a result, a total of 7410 sister pairs were included in the analysis. Breast cancer diagnosis after enrolment (i.e. incident cancers) was determined using self-reported and linked cancer registry data (ICD9 code 174 or ICD10 code C50); only the first incident invasive breast cancer diagnosis was considered. The affected women had an average age at diagnosis of 64 years. The 313-SNP PRS was calculated using published estimates of the OR per effect allele and effect allele frequency (Mavaddat et al., 2019) and scaled to have a population average risk of 1; see Spaeth et al. (2023) for more details. ICE FALCON analysis was conducted for the whole sample, and by age groups (<50 years, 50–60 years, ≥60 years). Both sisters of a pair were included in the analysis, and generalised estimating equation was used to account for the relatedness between sisters. Age was included in the models as a covariate. Association estimates were reported as the Odds ratio PER Adjusted standard deviation (OPERA) (Hopper, 2015). The causal effect of a PRS was calculated as $\log(\mathrm{OR}) = ([\beta_{\mathrm{relative}} - \beta'_{\mathrm{relative}}] - [\beta_{\mathrm{self}} - \beta'_{\mathrm{self}}])/\rho$, where ρ is the correlation in the PRS between sister pairs. Standard errors for the changes in regression coefficient estimates between models and for the causal effect were estimated using bootstrapping, as in Li et al. (2020).

2.3 | ICE CRISTAL analysis

ICE CRISTAL extends ICE FALCON by replacing $\mathrm{PRS}_{\mathrm{relative}}$ with $\mathrm{BC}_{\mathrm{relative}}$ in Models 2 and 3 above. $\mathrm{BC}_{\mathrm{relative}}$ is a component of a woman's breast cancer family history; therefore, $\mathrm{BC}_{\mathrm{relative}}$ can be further replaced with any measure of family history and then data for unrelated individuals can be used. That is, ICE CRISTAL estimates the marginal associations of a woman's breast cancer diagnosis with her own PRS, and with her family history, and then the conditional associations from fitting her PRS and her family history together, and finally considers the changes in the pairs of association estimates. Observing that the PRS association does not change after conditioning on family history, but that the family history association decreases after conditioning on PRS, is consistent with PRS having a causal effect on breast cancer. Familial confounding is consistent with both the PRS association and the family history association decreasing. Combinations of causation and familial confounding can occur, with the results are expected to a mixture of those arising from causation and those due to familial confounding.

2.4 | Applications of ICE CRISTAL analysis

We applied ICE CRISTAL to three studies. The first is a case-control study using the Australian Breast Cancer Family Registry (ABCFR), which recruited population-based breast cancer case families and age-matched control families and was designed to be enriched for breast cancer diagnosed at younger ages (Hopper et al., 1994, 1999; McCredie et al., 1998). A total of 750 breast cancer cases and 405 controls younger than 50 years old at diagnosis and recruitment, respectively, were included in this analysis. The average age at

10982272, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ggpi.22556 by National Health And Medical Research Council, Wiley Online Library on [02/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

diagnosis of cases was 38 years. The 77-SNP PRS was calculated using published estimates of the OR per effect allele and effect allele frequency (Mavaddat et al., 2015) and scaled to have a population average risk of 1, using the same method as above when applying ICE FALCON to the UK Biobank. Family history was used to calculate a familial risk score defined as the 5-year absolute risk of invasive breast cancer using the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA) (Antoniou et al., 2004, 2008). More details about the sampling, PRS and familial risk score calculations can be found in Dite et al. (2016). Logistic regression was used to fit the three ICE CRISTAL models, with age as a covariate. Associations were reported as the OPERAs of the PRS and familial risk score.

The second is a cohort study using the UK Biobank based on 200,195 women who were breast cancer free at enrolment. We considered the first invasive breast cancer diagnosis within 5 years of follow-up since baseline assessment. The sample included 197,057 unaffected women who had an average age of 57 years at baseline assessment, and 3138 affected women who had an average age of 58 years at baseline assessment and an average age at diagnosis of 61 years. The 77-SNP PRS and 313-SNP PRS were calculated as described above. Breast cancer family history in first-degree relatives (mother and/or sisters) was self-reported by the women at baseline assessment and analysed as a three-level categorical variable (no family history, one relative affected, two relatives affected). More details about the sampling and PRS calculation can be found in Spaeth et al. (2023). We used Cox regression to fit the three ICE CRISTAL models and estimated the hazard ratios per adjusted standard deviation of PRS and for family history, respectively.

The third is a case-cohort study nested in the Melbourne Collaborative Cohort Study (MCCS) of 24,469 women from Melbourne, Australia, aged between 27 and 76 years (99% were 40–69 years) at recruitment (Milne et al., 2017). The sample included 408 women diagnosed with the first invasive breast cancer within 10 years of follow-up and a random sample of 2783 women attending at baseline, of whom 93 were cases. The average age at diagnosis of cases was 68 years. Two 10-year absolute risk scores based on the 313-SNP PRS and family history were calculated using the BOADICEA risk model (Version 5) (Lee et al., 2019), respectively. More details about the sampling and risk score calculations can be found in Li et al. (2021). We used Cox regression to fit the three ICE CRISTAL models and estimated the hazard ratios per 10-year risk for the two risk scores.

2.5 | Further applications of ICE CRISTAL

We also examined the published data from essentially ICE CRISTAL analyses in the Breast Cancer Association Consortium (BCAC) and FinnGen. In the BCAC studies (Mavaddat et al., 2015, 2019), PRSs included the 77-SNP PRS and 313-SNP PRS, and family history was defined by having breast cancer in at least one first-degree relative. The 77-SNP PRS analysis involved 21,865 breast cancer cases and 15,830 controls (Mavaddat et al., 2015), and the 313-SNP PRS analysis involved 6787 ER-positive cancer cases and 17,351 controls and 4440 ER-negative cancer cases and 13,132 controls (Mavaddat et al., 2019).

In the FinnGen study (Mars et al., 2022), the PRS was genome-wide and consisted of 1,079,089 SNPs. There were three measures of family history: (1) breast cancer in at least one first-degree relative, (2) breast cancer in at least one second-degree relative and (3) breast cancer as the parental cause of death. The sample sizes were 886 cases and 14,395 controls, 881 cases and 18,092 controls and 8733 cases and 124,920 controls for the three measures of family history, respectively.

3 | RESULTS

3.1 | ICE FALCON results

Table 1 shows the results of ICE FALCON analyses applied to the UK Biobank sister pairs. Overall, a woman's breast cancer diagnosis was associated with her own PRS (OPERA = 1.51, 95% CI: 1.43, 1.60) and her sister's PRS (OPERA = 1.30, 95% CI: 1.22, 1.38). When the two PRSs were conditioned on each other (Model 3), the association with a woman's PRS remained unchanged, and the association with her sister's PRS attenuated 71% ($p = 10^{-16}$) to be at best marginally significant (OPERA = 1.08, 95% CI: 0.99, 1.17; $p = 0.07$). These results are consistent with, overall, the PRS having a causal effect on breast cancer, and the OPERA of the causal effect was estimated to be 1.40 (95% CI: 1.21, 1.63), accounting for 81% of the association between PRS and breast cancer.

When broken down by age at diagnosis and therefore age of controls, similar results were found for pairs aged 50–60 years and pairs aged ≥60 years, with the causal effect accounting for 75% and 83% of the association, respectively. For ages <50 years, however, a woman's breast cancer diagnosis was associated with her own PRS but not significantly with her sister's PRS (OPERA = 0.91, 95% CI: 0.54, 1.55) despite the PRS being naturally correlated within sister pairs. There was no evidence that

10982272, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ggpi.22556 by National Health And Medical Research Council, Wiley Online Library on [02/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 ICE FALCON results from the UK Biobank.[a]

Coefficient	Model 1 OPERA (95% CI)	Model 1 *p*	Model 2 OPERA (95% CI)	Model 2 *p*	Model 3 OPERA (95% CI)	Model 3 *p*	Change Estimate (SE)	Change *p*	Causal effect OPERA (95% CI)	Proportion of association due to causation (%)
Age at diagnosis < 50 years (871 pairs)										
β_{self}	1.34 (1.09, 1.63)	4.6×10^{-3}			1.44 (1.14, 1.81)	1.8×10^{-3}	0.073 (0.070)	0.30	—	—
β_{sister}			0.91 (0.54, 1.55)	0.62	0.76 (0.48, 1.20)	0.24	−0.190 (0.109)	0.08		
Age at diagnosis 50–60 years (1589 pairs)										
β_{self}	1.45 (1.27, 1.67)	1.0×10^{-7}			1.41 (1.16, 1.71)	5.1×10^{-4}	−0.031 (0.053)	0.56	1.32 (0.92, 2.39)	74.8
β_{sister}			1.28 (1.06, 1.55)	0.01	1.09 (0.83, 1.44)	0.53	−0.157 (0.066)	0.02		
Age at diagnosis ≥ 60 years (2036 pairs)										
β_{self}	1.58 (1.41, 1.78)	2.6×10^{-14}			1.55 (1.34, 1.79)	5.2×10^{-9}	−0.023 (0.032)	0.48	1.46 (1.13, 1.99)	82.9
β_{sister}			1.29 (1.15, 1.45)	2.6×10^{-5}	1.06 (0.91, 1.24)	0.46	−0.195 (0.041)	2.2×10^{-6}		
Whole sample (7410 pairs)										
β_{self}	1.51 (1.43, 1.60)	5.9×10^{-49}			1.47 (1.37, 1.57)	3.3×10^{-29}	−0.030 (0.017)	0.08	1.40 (1.21, 1.63)	81.3
β_{sister}			1.30 (1.22, 1.38)	6.2×10^{-17}	1.08 (0.99, 1.17)	0.07	−0.184 (0.002)	9.0×10^{-17}		

Abbreviations: CI, confidence interval; ICE FALCON, Inference about Causation from Examination of FAmiliaL CONfounding; OPERA, Odds ratio PER Adjusted standard deviation; SE, standard error.

[a]Notice that β_{self} is age-dependent.

10982272, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/gepi.22556 by National Health And Medical Research Council, Wiley Online Library on [02/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

the two PRS associations changed after conditioning on each other. These results are inconsistent with the PRS having a causal effect on breast cancer before ages 50 years.

3.2 | ICE CRISTAL results

Table 2 shows the results of ICE CRISTAL analyses of the three studies. In the ABCFR case-control study in which all breast cancer cases were diagnosed <50 years (average diagnosis age: 38 years) and all controls were also <50 years old, both the 77-SNP PRS association and family history association remained unchanged after conditioning on each other. These results are inconsistent with the 77-SNP PRS having a causal effect on breast cancer diagnosed before age 50 years, or explaining family history as a risk factor for breast cancers diagnosed before age 50 years.

In the UK Biobank study, where breast cancer cases had an average age at diagnosis of 61 years, both the 77-SNP PRS association and the 313-SNP PRS association remained unchanged after conditioning on family history, while the family history association attenuated 7%–10% after conditioning on the 77-SNP PRS and 11%–16% after conditioning on the 313-SNP PRS. These results are consistent with both the PRSs having causal effects on breast cancer diagnosed within 5 years of this older age range, with the 313-SNP PRS having a larger effect.

Results from the MCCS study, in which the breast cancer cases had an average age at diagnosis of 68 years, were consistent with the 313-SNP PRS having a causal effect on breast cancers diagnosed within 10 years of this older age range, as the PRS association remained unchanged after conditioning on family history, and the family history association attenuated 70% after conditioning on PRS.

Published results from the BCAC and FinnGen studies (Mars et al., 2022; Mavaddat et al., 2015, 2019) were also consistent with PRS having a causal effect (Table 2). In the BCAC study, the family history association attenuated 13% after conditioning on the 77-SNP PRS. The attenuation was only 5% at ages <40 years, but greater at older ages. As for the 313-SNP PRS, the family history associations with breast cancer subtypes attenuated after conditioning on it, with a greater attenuation for ER-positive breast cancer. In the FinnGen study, family history associations attenuated 16%–21% after conditioning on the genome-wide PRS, with the association for family history in a second-degree relative having the smallest attenuation. The PRS association remained unchanged after conditioning on having a family history in a second-degree relative or breast cancer as parental cause of death and attenuated 6% after conditioning on having a family history in a first-degree relative.

4 | DISCUSSION

Genetic epidemiology is about how genetic and non-genetic factors combine to influence health and disease, realising that familial associations might not necessarily due to genes, and that genetic associations might not necessarily be causal. Genetic epidemiology traditionally evolved from considering families, or at least sets of related individuals, as the unit of study and not isolated individuals alone. This study has taken genetic epidemiology back to its roots and used the experiment of opportunity arising from family studies to address causation, the Holy Grail of science, by applying the novel ICE FALCON and ICE CRISTAL concepts.

We found consistent evidence that the breast cancer PRSs could be picking up signals from breast cancer causes when the data were pooled over all ages at diagnosis, though not necessarily for breast cancers diagnosed at a young age such as before age 50 years. The causality might also be due to the genetic variants involved in the PRS being in linkage disequilibrium with truly causal variants and not necessarily being casual themselves. Evidence of causation came from both the ICE FALCON analyses of sibling pairs, and the individual-specific ICE CRISTAL analyses of the family history association attenuation magnitude after conditioning on PRS, which increased with the number of variants included in the PRS: 77-SNP PRS versus 313-SNP PRS in the UK Biobank, and 313-SNP PRS versus genome-wide PRS in the FinnGen (Mars et al., 2022); a PRS consisting of more variants appears to be capturing the effects of more causal variants.

Leaving aside the paucity of evidence for the PRSs causing breast cancer diagnosed before age 50 years (see below), the causal effects being captured by the PRS are more evident and stronger for breast cancers diagnosed at older ages; ICE FALCON analyses found that the proportion of association due to causality increased with age and the ICE CRISTAL analyses found that the magnitude of family history attenuation after conditioning on PRS increased with age.

An important observation was that (1) a woman's breast cancer was not associated with her sister's PRS at ages <50 years in the UK Biobank study, (2) the family history association did not change after conditioning on

10982272, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ggpi.22556 by National Health And Medical Research Council, Wiley Online Library on [02/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 ICE CRISTAL results from three cohorts and published data.

Predictors	Model 1 OR or HR	Model 2 OR or HR	Model 3 OR or HR	Change (%)
Australian Breast Cancer Family Registry (age at diagnosis and age of controls < 50 years)				
PRS (77 SNPs)	1.46		1.46	0
Family history (BOADICEA 5-year risk)		1.80	1.80	0
UK Biobank (Average age at diagnosis = 61 years)				
Analysis 1				
PRS (77 SNPs)	1.62		1.61	1.2
Family history (one first-degree relative affected)		1.58	1.51	9.9
Family history (two first-degree relatives affected)		2.56	2.39	7.3
Analysis 2				
PRS (313 SNPs)	1.53		1.52	1.5
Family history (one first-degree relative affected)		1.58	1.47	15.8
Family history (two first-degree relatives affected)		2.56	2.31	10.9
Melbourne Collaborative Cohort Study (average age at diagnosis = 68 years)				
PRS (313 SNPs)	2.76		2.69	2.5
Family history (BOADICEA 10-year risk)		1.84	1.20	70.0
Breast Cancer Association Consortium				
PRS (77 SNPs)	1.55		—	
Family history (first-degree relatives)				
<40 years		2.90	2.76	4.6
40–60 years		1.88	1.72	14.1
≥60 years		1.63	1.53	13.0
Whole sample		1.81	1.68	12.6
Breast Cancer Association Consortium				
ER-negative breast cancer				
PRS (313 SNPs)	1.44		—	
Family history (first-degree relatives)		1.66	1.56	12.3
ER-positive breast cancer				
PRS (313 SNPs)	1.67		—	
Family history (first-degree relatives)		1.59	1.44	21.4
FinnGen				
Analysis 1				
PRS (1,079,089 SNPs)	1.71		1.66	5.5
Family history (first-degree relatives)		2.36	1.99	19.9

10982272, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/gepi.22556 by National Health And Medical Research Council, Wiley Online Library on [02/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 (Continued)

Predictors	Model 1 OR or HR	Model 2 OR or HR	Model 3 OR or HR	Change (%)
Analysis 2				
PRS (1,079,089 SNPs)	1.84		1.83	0.9
Family history (second-degree relatives)		1.47	1.38	16.4
Analysis 3				
PRS (1,079,089 SNPs)	1.76		1.76	0
Family history (parental cause of death)		2.02	1.74	21.2

Abbreviations: BOADICEA, Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm; HR, hazard ratio; ICE CRISTAL, Inference about Causation from Examining Changes in Regression coefficients and Innovative STatistical AnaLyses; OR, odds ratio; PRS, polygenic risk score; SNP, single-nucleotide polymorphism.

PRS at ages <50 years in the ABCFR study and (3) the family history association only decreased 5% after conditioning on PRS at ages <40 years in the BCAC study. These observations imply that the PRSs explain minimal familial aggregation before menopause and at young ages, when the familial relative risks are greater, and are consistent with previous findings from segregation analyses (Li et al., 2022). This could be mostly due to the breast cancer familial risk, or familial variance (Hopper et al., 2023), decreasing with age (Collaborative Group on Hormonal Factors in Breast, 2001; Li et al., 2022; Wang et al., 2023) But the effect size of the within-woman PRS association stays relatively stable with age (Li & Hopper, 2021; Li, 2021; Mavaddat et al., 2015, 2019), an intriguing observation given that risk associations should in general decrease with age as those at higher risk are preferentially 'eliminated' from older control groups. For colorectal and prostate cancer, the magnitudes of the PRS risk associations decrease substantially with age (Archambault et al., 2020; Li, 2021; Li & Hopper, 2021; Schaid et al., 2022; Thomas et al., 2020).

There is evidence that the PRS associations are confounded by nongenetic familial factors shared within nuclear families. This comes from observations reported by the FinnGen study that (1) the PRS association decreased 6% after conditioning on family history in first-degree relatives, (2) there was no attenuation in the PRS association after conditioning on family history in second-degree relatives and (3) the family history association decreased more after conditioning on a PRS when it was defined in terms of first-degree rather than second-degree relatives. This is further supported by the ICE FALCON analyses which found that there was a marginally statistically significant ($p = 0.08$) change in a woman's own PRS association after conditioning on her sister's PRS, and the sister's PRS association was marginally statistically significant ($p = 0.07$) in Model 3. As discussed in the Introduction, the familial confounding could be due to factors such as assortative mating, population stratification and indirect genetic effects (Brumpton et al., 2020).

There is evidence that there might be nongenetic factors contributing to breast cancer familial aggregation. Under the Variance of Age-specific Log Incidence Decomposition (VALID) model (Hopper et al., 2023), the familial variance in breast cancer log(incidence) is given by $\sigma^2 = \log(\mathrm{FRR})/r$, where FRR is the familial risk ratio for a given set of relatives and r is the correlation between those relatives in the familial causes. Assuming a model in which all of the breast cancer familial variance is due to genetic factors, the FRR for first-degree relatives (in which case $r = 0.5$) in FinnGen predicts the familial variance to be 1.7, larger than the 1.5 based on the FRR for second-degree relatives (in which case $r = 0.25$). The excess familial variance based on first-degree relatives implies the existence of nongenetic familial factors. This is also suggested by analysis of twin pairs in the Nordic Twin Study which found evidence for nongenetic factors shared by twins, especially at younger ages (Hopper et al., 2023; Möller et al., 2016).

The study of Mars et al. (2022) using the FinnGen data investigated many other common diseases including other cancers, eye diseases, chronic kidney disease, and cardiometabolic, psychiatric, musculoskeletal, neurological, endocrine and allergic disorders. Consistent across the investigated diseases, they found that (1) there is evidence for nongenetic causes of disease familial

10982272, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ggpi.22556 by National Health And Medical Research Council, Wiley Online Library on [02/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

aggregation, (2) the PRSs capture a substantial causal effect and (3) some of the nongenetic familial causes of diseases are confounded with PRS; see Li and Hopper (2023) for more discussion of the study implications.

Our study has demonstrated the unique value of the family-based approaches of ICE FALCON and ICE CRISTAL in providing evidence for both causation and familial confounding. Unlike Mendelian randomisation, these two methods can be used to investigate the causal effect of PRS and are in general more powerful and flexible. Both methods use the measured relative's exposure or family history as a proxy for true instrumental variables that are not observed (i.e. familial causes that are truly specific to the exposure, not just assumed to be as in Mendelian randomisation). Not limited to measured 'weak' genetic variants as instrumental variables, the two methods can investigate all measured exposures, not just those for which there is substantial genetic knowledge as for Mendelian randomisation. The two methods can also make causal inference in the presence of familial confounding (in which the proxy instrumental variable is associated with the outcome through pathways not involving the exposure).

We have found here that the ICE FALCON and ICE CRISTAL methods both supported and complemented each other. The proxy instrumental variable is stronger in ICE FALCON than in ICE CRISTAL, so is more powerful because the relative's exposure is more correlated with a person's exposure than family history (i.e. the relative's outcome). On the other hand, ICE CRISTAL analysis is more widely applicable and can potentially be applied to more and larger sample sizes, given only a small proportion of studies collect data from relatives but most collect family history data. The two methods can be used to attempt to cross-validate each other's findings using the same study, as we have shown.

In conclusion, the investigated breast cancer PRSs capture causal effects, at least for postmenopausal disease, and the causal effect is more evident and stronger at older ages. These current PRSs, however, cause minimal heritability of breast cancer at younger ages. There is also evidence for nongenetic factors shared by first-degree relatives that also explain familial aggregation of breast cancer. Familial associations are not necessarily due to genes, and genetic associations are not necessarily causal.

ACKNOWLEDGEMENTS

The UK Biobank resource used in this study was accessed under Application Number 47401. This study is supported by National Health and Medical Research Council (NHMRC) grant GNT2024867. The ABCFR was supported in Australia by NHMRC, the New South Wales Cancer Council, the Victorian Health Promotion Foundation, the Victorian Breast Cancer Research Consortium, Cancer Australia, and the National Breast Cancer Foundation. The six sites of the Breast Cancer Family Registry (BCFR) were supported by grant UM1 CA164920 from the U.S. National Cancer Institute. The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centres in the BCFR, nor does mention of trade names, commercial products, or organisations imply endorsement by the U.S. Government or the BCFR. MCCS cohort recruitment was funded by Cancer Council Victoria and VicHealth. The MCCS was further supported by NHMRC grants 209057, 396414, 1074383, and 1129136, and ongoing follow-up and data management have been funded by Cancer Council Victoria since 1995. Cases and their vital status were ascertained through the Victorian Cancer Registry and the Australian Institute of Health and Welfare, including the National Death Index and the Australian Cancer Database. S. L. is an NHMRC Emgerging Leadership Fellow (GNT2017373). M. C. S. is an NHMRC Leadership Fellow. J. L. H. is a Dame Kate Campbell Professorial Fellow of the University of Melbourne. Open access publishing facilitated by The University of Melbourne, as part of the Wiley - The University of Melbourne agreement via the Council of Australian University Librarians.

CONFLICT OF INTEREST STATEMENT

G. S. D. is an employee of Genetic Technologies Limited. The remaining authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Shuai Li http://orcid.org/0000-0002-8696-8594
Zhoufeng Ye https://orcid.org/0000-0002-0884-4246

REFERENCES

Antoniou, A. C., Cunningham, A. P., Peto, J., Evans, D. G., Lalloo, F., Narod, S. A., Risch, H. A., Eyfjord, J. E., Hopper, J. L., Southey, M. C., Olsson, H., Johannsson, O., Borg, A., Passini, B., Radice, P., Manoukian, S., Eccles, D. M., Tang, N., Olah, E., ... Easton, D. F. (2008). The BOADICEA model of genetic susceptibility to breast and ovarian cancers:

10982272, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ggpi.22556 by National Health And Medical Research Council, Wiley Online Library on [02/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Updates and extensions. *British Journal of Cancer*, *98*(8), 1457–1466. https://doi.org/10.1038/sj.bjc.6604305

Antoniou, A. C., Pharoah, P. P. D., Smith, P., & Easton, D. F. (2004). The BOADICEA model of genetic susceptibility to breast and ovarian cancer. *British Journal of Cancer*, *91*(8), 1580–1590. https://doi.org/10.1038/sj.bjc.6602175

Archambault, A. N., Su, Y. R., Jeon, J., Thomas, M., Lin, Y., Conti, D. V., Win, A. K., Sakoda, L. C., Lansdorp-Vogelaar, I., Peterse, E. F. P., Zauber, A. G., Duggan, D., Holowatyj, A. N., Huyghe, J. R., Brenner, H., Cotterchio, M., Bézieau, S., Schmit, S. L., Edlund, C. K., ... Hayes, R. B. (2020). Cumulative burden of colorectal cancer-associated genetic variants is more strongly associated with early-onset vs late-onset cancer. *Gastroenterology*, *158*(5), 1274–1286.e12. https://doi.org/10.1053/j.gastro.2019.12.012

Brumpton, B., Sanderson, E., Heilbron, K., Hartwig, F. P., Harrison, S., Vie, G. Å., Cho, Y., Howe, L. D., Hughes, A., Boomsma, D. I., Havdahl, A., Hopper, J., Neale, M., Nivard, M. G., Pedersen, N. L., Reynolds, C. A., Tucker-Drob, E. M., Grotzinger, A., Howe, L., ... Davies, N. M. (2020). Avoiding dynastic, assortative mating, and population stratification biases in Mendelian randomization through within-family analyses. *Nature Communications*, *11*(1), 3519. https://doi.org/10.1038/s41467-020-17117-4

Bui, M., Bjørnerem, Å., Ghasem-Zadeh, A., Dite, G. S., Hopper, J. L., & Seeman, E. (2013). Architecture of cortical bone determines in part its remodelling and structural decay. *Bone*, *55*(2), 353–358. https://doi.org/10.1016/j.bone.2013.04.020

Bui, M., Zebaze, R., Li, S., Hopper, J. L., & Bjørnerem, Å. (2020). Are the relationships of lean mass and fat mass with bone microarchitecture causal or due to familial confounders? A novel study of adult female twin pairs. *JBMR Plus*, *4*(9), e10386. https://doi.org/10.1002/jbm4.10386

Collaborative Group on Hormonal Factors in Breast. (2001). Familial breast cancer: Collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. *The Lancet*, *358*(9291), 1389–1399. https://doi.org/10.1016/S0140-6736(01)06524-2

Dite, G. S., Gurrin, L. C., Byrnes, G. B., Stone, J., Gunasekara, A., McCredie, M. R. E., English, D. R., Giles, G. G., Cawson, J., Hegele, R. A., Chiarelli, A. M., Yaffe, M. J., Boyd, N. F., & Hopper, J. L. (2008). Predictors of mammographic density: Insights gained from a novel regression analysis of a twin study. *Cancer Epidemiology, Biomarkers & Prevention*, *17*(12), 3474–3481. https://doi.org/10.1158/1055-9965.EPI-07-2636

Dite, G. S., MacInnis, R. J., Bickerstaffe, A., Dowty, J. G., Allman, R., Apicella, C., Milne, R. L., Tsimiklis, H., Phillips, K. A., Giles, G. G., Terry, M. B., Southey, M. C., & Hopper, J. L. (2016). Breast cancer risk prediction using clinical models and 77 independent risk-associated SNPs for women aged under 50 years: Australian Breast Cancer Family Registry. *Cancer Epidemiology, Biomarkers & Prevention*, *25*(2), 359–365. https://doi.org/10.1158/1055-9965.epi-15-0838

Dongmeng, W., Yu'e, X., Wenjing, G., Ke, Z., Jun, L., Canqing, Y., Shengfeng, W., Tao, H., Dianjianyi, S., Chunxiao, L., Yuanjie, P., Zengchang, P., Min, Y., Hua, W., Xianping, W., Zhong, D., Fan, W., Guohong, J., Xiaojie, W., ... Liming, L. (2022). Heritability of tea drinking and its relationship with cigarette smoking in the Chinese male adult twins. *Addiction Biology*, *27*(2), e13129. https://doi.org/10.1111/adb.13129

Hong, X., Miao, K., Cao, W., Lv, J., Yu, C., Huang, T., Sun, D., Liao, C., Pang, Y., Pang, Z., Yu, M., Wang, H., Wu, X., Liu, Y., Gao, W., & Li, L. (2023). Association between DNA methylation and blood pressure: A 5-year longitudinal twin study. *Hypertension*, *80*(1), 169–181. https://doi.org/10.1161/hypertensionaha.122.19953

Hopper, J. L. (2015). Odds per adjusted standard deviation: Comparing strengths of associations for risk factors measured on different scales and across diseases and populations. *American Journal of Epidemiology*, *182*(10), 863–867. https://doi.org/10.1093/aje/kwv193

Hopper, J. L., Bui, Q. M., Erbas, B., Matheson, M. C., Gurrin, L. C., Burgess, J. A., Lowe, A. J., Jenkins, M. A., Abramson, M. J., Walters, E. H., Giles, G. G., & Dharmage, S. C. (2012). Does eczema in infancy cause hay fever, asthma, or both in childhood? Insights from a novel regression model of sibling data. *Journal of Allergy and Clinical Immunology*, *130*(5), 1117–1122.e1. https://doi.org/10.1016/j.jaci.2012.08.003

Hopper, J. L., Chenevix-Trench, G., Jolley, D. J., Dite, G. S., Jenkins, M. A., Venter, D. J., McCredie, M. R. E., & Giles, G. G. (1999). Design and analysis issues in a population-based, case-control-family study of the genetic epidemiology of breast cancer and the Co-Operative Family Registry for Breast Cancer Studies (CFRBCS). *JNCI Monographs*, *1999*, 95–100. https://doi.org/10.1093/oxfordjournals.jncimonographs.a024232

Hopper, J. L., Dowty, J. G., Nguyen, T. L., Li, S., Dite, G. S., MacInnis, R. J., Makalic, E., Schmidt, D. F., Bui, M., Stone, J., Sung, J., Jenkins, M. A., Giles, G. G., Southey, M. C., & Mathews, J. D. (2023). Variance of age-specific log incidence decomposition (VALID): A unifying model of measured and unmeasured genetic and non-genetic risks. *International Journal of Epidemiology*, *52*, 1557–1568. https://doi.org/10.1093/ije/dyad086

Hopper, J. L., Giles, G. G., McCredie, M. R. E., & Boyle, P. (1994). Background, rationale and protocol for a case-control-family study of breast cancer. *The Breast*, *3*(2), 79–86. https://doi.org/10.1016/0960-9776(94)90003-5

Lee, A., Mavaddat, N., Wilcox, A. N., Cunningham, A. P., Carver, T., Hartley, S., Babb de Villiers, C., Izquierdo, A., Simard, J., Schmidt, M. K., Walter, F. M., Chatterjee, N., Garcia-Closas, M., Tischkowitz, M., Pharoah, P., Easton, D. F., & Antoniou, A. C. (2019). BOADICEA: A comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. *Genetics in Medicine*, *21*(8), 1708–1718. https://doi.org/10.1038/s41436-018-0406-9

Li, S. (2021). Negative age-dependence of the polygenic risk score gradient for colorectal cancer. *Gastroenterology*, *160*(6), 2214–2215. https://doi.org/10.1053/j.gastro.2020.09.064

Li, S., Bui, M., & Hopper, J. L. (2020). Inference about causation from examination of familial confounding (ICE FALCON): A model for assessing causation analogous to Mendelian randomization. *International Journal of Epidemiology*, *49*(4), 1259–1269. https://doi.org/10.1093/ije/dyaa065

Li, S., & Hopper, J. L. (2021). Age dependency of the polygenic risk score for colorectal cancer. *The American Journal of Human*

10982272, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ggpi.22556 by National Health And Medical Research Council, Wiley Online Library on [02/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Genetics, *108*(3), 525–526. https://doi.org/10.1016/j.ajhg.2021.02.002

Li, S., & Hopper, J. L. (2023). Implications of family history and polygenic risk scores for causation. *The American Journal of Human Genetics*, *110*(7), 1221–1223. https://doi.org/10.1016/j.ajhg.2023.05.015

Li, S., MacInnis, R. J., Lee, A., Nguyen-Dumont, T., Dorling, L., Carvalho, S., Dite, G. S., Shah, M., Luccarini, C., Wang, Q., Milne, R. L., Jenkins, M. A., Giles, G. G., Dunning, A. M., Pharoah, P. D. P., Southey, M. C., Easton, D. F., Hopper, J. L., & Antoniou, A. C. (2022). Segregation analysis of 17,425 population-based breast cancer families: Evidence for genetic susceptibility and risk prediction. *The American Journal of Human Genetics*, *109*(10), 1777–1788. https://doi.org/10.1016/j.ajhg.2022.09.006

Li, S., Wong, E. M., Bui, M., Nguyen, T. L., Joo, J. H. E., Stone, J., Dite, G. S., Dugué, P. A., Milne, R. L., Giles, G. G., Saffery, R., Southey, M. C., & Hopper, J. L. (2019). Inference about causation between body mass index and DNA methylation in blood from a twin family study. *International Journal of Obesity*, *43*(2), 243–252. https://doi.org/10.1038/s41366-018-0103-4

Li, S., Wong, E. M., Bui, M., Nguyen, T. L., Joo, J. H. E., Stone, J., Dite, G. S., Giles, G. G., Saffery, R., Southey, M. C., & Hopper, J. L. (2018). Causal effect of smoking on DNA methylation in peripheral blood: A twin and family study. *Clinical Epigenetics*, *10*(1), 18. https://doi.org/10.1186/s13148-018-0452-9

Li, S. X., Milne, R. L., Nguyen-Dumont, T., Wang, X., English, D. R., Giles, G. G., Southey, M. C., Antoniou, A. C., Lee, A., Li, S., Winship, I., Hopper, J. L., Terry, M. B., & MacInnis, R. J. (2021). Prospective evaluation of the addition of polygenic risk scores to breast cancer risk models. *JNCI Cancer Spectrum*, *5*(3), pkab021. https://doi.org/10.1093/jncics/pkab021

Mars, N., Lindbohm, J. V., Della Briotta Parolo, P., Widén, E., Kaprio, J., Palotie, A., & Ripatti, S. (2022). Systematic comparison of family history and polygenic risk across 24 common diseases. *The American Journal of Human Genetics*, *109*(12), 2152–2162. https://doi.org/10.1016/j.ajhg.2022.10.009

Martin, L. J., Melnichouk, O., Guo, H., Chiarelli, A. M., Hislop, T. G., Yaffe, M. J., Minkin, S., Hopper, J. L., & Boyd, N. F. (2010). Family history, mammographic density, and risk of breast cancer. *Cancer Epidemiology, Biomarkers & Prevention*, *19*(2), 456–463. https://doi.org/10.1158/1055-9965.Epi-09-0881

Mavaddat, N., Michailidou, K., Dennis, J., Lush, M., Fachal, L., Lee, A., Tyrer, J. P., Chen, T. H., Wang, Q., Bolla, M. K., Yang, X., Adank, M. A., Ahearn, T., Aittomäki, K., Allen, J., Andrulis, I. L., Anton-Culver, H., Antonenkova, N. N., Arndt, V., … Easton, D. F. (2019). Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. *The American Journal of Human Genetics*, *104*(1), 21–34. https://doi.org/10.1016/j.ajhg.2018.11.002

Mavaddat, N., Pharoah, P. D. P., Michailidou, K., Tyrer, J., Brook, M. N., Bolla, M. K., Wang, Q., Dennis, J., Dunning, A. M., Shah, M., Luben, R., Brown, J., Bojesen, S. E., Nordestgaard, B. G., Nielsen, S. F., Flyger, H., Czene, K., Darabi, H., Eriksson, M., … Garcia-Closas, M. (2015). Prediction of breast cancer risk based on profiling with common genetic variants. *JNCI: Journal of the National Cancer Institute*, *107*(5), djv036. https://doi.org/10.1093/jnci/djv036

McCredie, M. R. E., Dite, G. S., Giles, G. G., & Hopper, J. L. (1998). Breast cancer in Australian women under the age of 40. *Cancer Causes and Control*, *9*(2), 189–198. https://doi.org/10.1023/a:1008886328352

Milne, R. L., Fletcher, A. S., MacInnis, R. J., Hodge, A. M., Hopkins, A. H., Bassett, J. K., Bruinsma, F. J., Lynch, B. M., Dugué, P. A., Jayasekara, H., Brinkman, M. T., Popowski, L. V., Baglietto, L., Severi, G., O'Dea, K., Hopper, J. L., Southey, M. C., English, D. R., & Giles, G. G. (2017). Cohort profile: The Melbourne Collaborative Cohort Study (Health 2020). *International Journal of Epidemiology*, *46*(6), 1757–1757i. https://doi.org/10.1093/ije/dyx085

Möller, S., Mucci, L. A., Harris, J. R., Scheike, T., Holst, K., Halekoh, U., Adami, H. O., Czene, K., Christensen, K., Holm, N. V., Pukkala, E., Skytthe, A., Kaprio, J., & Hjelmborg, J. B. (2016). The heritability of breast cancer among women in the nordic twin study of cancer. *Cancer Epidemiology, Biomarkers & Prevention*, *25*(1), 145–150. https://doi.org/10.1158/1055-9965.Epi-15-0913

Nissen, F. I., Esser, V. F. C., Bui, M., Li, S., Hopper, J. L., Bjørnerem, Å., & Hansen, A. K. (2023). Is there a causal relationship between physical activity and bone microarchitecture? A study of adult female twin pairs. *Journal of Bone and Mineral Research*, *38*, 951–957. https://doi.org/10.1002/jbmr.4826

Schaid, D. J., Sinnwell, J. P., Batzler, A., & McDonnell, S. K. (2022). Polygenic risk for prostate cancer: Decreasing relative risk with age but little impact on absolute risk. *The American Journal of Human Genetics*, *109*(5), 900–908. https://doi.org/10.1016/j.ajhg.2022.03.008

Spaeth, E. L., Dite, G. S., Hopper, J. L., & Allman, R. (2023). Validation of an abridged breast cancer risk prediction model for the general population. *Cancer Prevention Research*, *16*(5), 281–291. https://doi.org/10.1158/1940-6207.Capr-22-0460

Stone, J., Dite, G. S., Giles, G. G., Cawson, J., English, D. R., & Hopper, J. L. (2012). Inference about causation from examination of familial confounding: Application to longitudinal twin data on mammographic density measures that predict breast cancer risk. *Cancer Epidemiology, Biomarkers & Prevention*, *21*(7), 1149–1155. https://doi.org/10.1158/1055-9965.EPI-12-0051

Thomas, M., Sakoda, L. C., Hoffmeister, M., Rosenthal, E. A., Lee, J. K., van Duijnhoven, F. J. B., Platz, E. A., Wu, A. H., Dampier, C. H., de la Chapelle, A., Wolk, A., Joshi, A. D., Burnett-Hartman, A., Gsur, A., Lindblom, A., Castells, A., Win, A. K., Namjou, B., Van Guelpen, B., … Hsu, L. (2020). Genome-wide modeling of polygenic risk score in colorectal cancer risk. *The American Journal of Human Genetics*, *107*(3), 432–444. https://doi.org/10.1016/j.ajhg.2020.07.006

Tian, H., Qiao, H., Han, F., Kong, X., Zhu, S., Xing, F., Duan, H., Li, W., Wang, W., Zhang, D., & Wu, Y. (2023). Genome-wide DNA methylation analysis of body composition in Chinese monozygotic twins. *European Journal of Clinical Investigation*, *53*, e14055. https://doi.org/10.1111/eci.14055

Wang, H., MacInnis, R. J., & Li, S. (2023). Family history and breast cancer risk for Asian women: A systematic review and

10982272, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ggpi.22556 by National Health And Medical Research Council, Wiley Online Library on [02/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

meta-analysis. *BMC Medicine*, *21*(1), 239. https://doi.org/10.1186/s12916-023-02950-3

Wu, Y., Tian, H., Wang, W., Li, W., Duan, H., & Zhang, D. (2022). DNA methylation and waist-to-hip ratio: An epigenome-wide association study in Chinese monozygotic twins. *Journal of Endocrinological Investigation*, *45*(12), 2365–2376. https://doi.org/10.1007/s40618-022-01878-4

Zheng, K., Gao, W., Cao, W., Lv, J., Yu, C., Wang, S., Huang, T., Sun, D., Liao, C., Pang, Y., Pang, Z., Yu, M., Wang, H., Wu, X., Dong, Z., Wu, F., Jiang, G., Wang, X., Liu, Y., … Li, L. (2022). Education, income, and obesity: A nationwide Chinese twin study. *Obesity*, *30*(4), 931–942. https://doi.org/10.1002/oby.23390

How to cite this article: Li, S., Dite, G. S., MacInnis, R. J., Bui, M., Nguyen, T. L., Esser, V. F. C., Ye, Z., Dowty, J. G., Makalic, E., Sung, J., Giles, G. G., Southey, M. C., & Hopper, J. L. (2024). Causation and familial confounding as explanations for the associations of polygenic risk scores with breast cancer: Evidence from innovative ICE FALCON and ICE CRISTAL analyses. *Genetic Epidemiology*, 1–13. https://doi.org/10.1002/gepi.22556

10982272, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/gepi.22556 by National Health And Medical Research Council, Wiley Online Library on [02/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License