



Breast cancer risk factors and their effects on survival a Mendelian randomisation study

Escala-Garcia, Maria; Morra, Anna; Canisius, Sander; Chang-Claude, Jenny; Kar, Siddhartha; Zheng, Wei; Bojesen, Stig E; Easton, Doug; Pharoah, Paul D P; Schmidt, Marjanka K

Published in:
BMC Medicine

DOI:
[10.1186/s12916-020-01797-2](https://doi.org/10.1186/s12916-020-01797-2)

Publication date:
2020

Document version
Publisher's PDF, also known as Version of record

Document license:
[CC BY](#)

Citation for published version (APA):
Escala-Garcia, M., Morra, A., Canisius, S., Chang-Claude, J., Kar, S., Zheng, W., Bojesen, S. E., Easton, D., Pharoah, P. D. P., & Schmidt, M. K. (2020). Breast cancer risk factors and their effects on survival: a Mendelian randomisation study. *BMC Medicine*, 18, [327]. <https://doi.org/10.1186/s12916-020-01797-2>

RESEARCH ARTICLE

Open Access

Breast cancer risk factors and their effects on survival: a Mendelian randomisation study



Maria Escala-Garcia¹, Anna Morra¹, Sander Canisius^{1,2}, Jenny Chang-Claude^{3,4}, Siddhartha Kar^{5,6}, Wei Zheng⁷, Stig E. Bojesen^{8,9,10}, Doug Easton^{11,12}, Paul D. P. Pharoah^{11,12} and Marjanka K. Schmidt^{1,13*} 

Abstract

Background: Observational studies have investigated the association of risk factors with breast cancer prognosis. However, the results have been conflicting and it has been challenging to establish causality due to potential residual confounding. Using a Mendelian randomisation (MR) approach, we aimed to examine the potential causal association between breast cancer-specific survival and nine established risk factors for breast cancer: alcohol consumption, body mass index, height, physical activity, mammographic density, age at menarche or menopause, smoking, and type 2 diabetes mellitus (T2DM).

Methods: We conducted a two-sample MR analysis on data from the Breast Cancer Association Consortium (BCAC) and risk factor summary estimates from the GWAS Catalog. The BCAC data included 86,627 female patients of European ancestry with 7054 breast cancer-specific deaths during 15 years of follow-up. Of these, 59,378 were estrogen receptor (ER)-positive and 13,692 were ER-negative breast cancer patients. For the significant association, we used sensitivity analyses and a multivariable MR model. All risk factor associations were also examined in a model adjusted by other prognostic factors.

Results: Increased genetic liability to T2DM was significantly associated with worse breast cancer-specific survival (hazard ratio [HR] = 1.10, 95% confidence interval [CI] = 1.03–1.17, *P* value [*P*] = 0.003). There were no significant associations after multiple testing correction for any of the risk factors in the ER-status subtypes. For the reported significant association with T2DM, the sensitivity analyses did not show evidence for violation of the MR assumptions nor that the association was due to increased BMI. The association remained significant when adjusting by other prognostic factors.

Conclusions: This extensive MR analysis suggests that T2DM may be causally associated with worse breast cancer-specific survival and therefore that treating T2DM may improve prognosis.

Keywords: Breast cancer survival, Breast cancer risk factors, Lifestyle, Mendelian randomisation, GWAS Catalog

* Correspondence: mk.schmidt@nki.nl

¹Division of Molecular Pathology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands

¹³Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Breast cancer is a heterogeneous disease with a broad variation in prognosis [1]. Providing a precise prognostication for breast cancer patients is important in order to inform them accurately about the course of the disease and to allocate them to the right treatment [2]. To date, most commonly used prognostic factors relate to tumour characteristics and the extent of the disease at the time of diagnosis [2]. Many observational studies have evaluated the association of breast cancer risk and survival with other patient characteristics and lifestyle-related risk factors [3–5]. However, due to their observational nature, it is difficult for these studies to establish causation. Understanding whether or not the association between breast cancer survival and risk factors is causal might influence strategies to improve survival in breast cancer patients. In theory, randomised control trials (RCTs) provide a reliable method to evaluate the causal relationship between risk factors and survival [6, 7], but they are often not feasible as they can be prohibitively expensive, time-consuming, and even unethical. If an RCT cannot be performed to assess the causal effect between a risk factor and the outcome of interest, methods using instrumental variables may be an alternative.

Mendelian randomisation (MR) is a popular analytical method that uses genetic variants as instrumental variables (i.e. genetic instruments). This methodology uses a genetic predictor for the risk factor. Because of the natural randomisation of alleles during meiosis, this genetic predictor will be independently distributed across a population. Theoretically, therefore, this genetic instrument is not affected by potential environmental confounding factors or by disease status. MR rests on three basic assumptions: (1) genetic variants are associated with the risk factor (relevance assumption), (2) those genetic variants are not associated with any known or unknown confounders (independence assumption), and (3) the genetic variants affect the outcome only through the risk factor (exclusion restriction assumption) [8]. Using a genetic score that combines multiple variants explaining a large *R*-squared of the risk factor can help reducing the probability of violating the first MR assumption and providing more powerful MR analyses. The third assumption is also known as independence from horizontal pleiotropy, which occurs when the genetic variants influence the outcome by means of other pathways independently of the risk factor [8]. Several methods and sensitivity tests exist to assess these assumptions [9].

In this study, we used MR analysis to evaluate the causal relationships between breast cancer-specific survival and nine established risk factors for breast cancer: alcohol consumption, body mass index (BMI), height, mammographic density, menarche (age at onset),

menopause (age at onset), physical activity, smoking, and type 2 diabetes mellitus (T2DM). Observational studies have provided evidence for the potential association of these risk factors and breast cancer survival, sometimes with conflicting results.

A population-based prospective study found that smoking before or after breast cancer diagnosis is associated with worse breast cancer survival [10]. Another meta-analysis of cohort studies concluded that current smoking is associated with worse breast cancer-specific survival compared to never smoking in breast cancer patients [11]. Obesity (BMI of ≥ 30.0) has been associated with worse breast cancer survival in a meta-analysis and systematic review [12]. In another review, obesity was associated with worse breast cancer prognosis for women of all ages [13]. For T2DM, a retrospective study of breast cancer patients found that diabetes was independently associated with poorer breast cancer prognosis [14]. In a population-based study, breast cancer-specific mortality was higher among women with diabetes compared to non-diabetic patients [15]. In relation to menstrual risk factors, a population-based study showed that early age at menarche was significantly associated with poorer survival but age at menopause did not have a significant impact [16]. The relationship between mammographic density and breast cancer survival has been studied in several cohort studies, but results have been inconclusive [17–19]. For other factors such as physical activity, the evidence is also not clear: in an RCT with an 8-year follow-up, no significant difference in disease-free survival was found between an exercise group and a usual care group [20]. To date, there is no evidence for an association between height or post-diagnosis alcohol consumption and breast cancer survival [21].

Our hypothesis was that some of these risk factors, for which there is evidence of an association with breast cancer survival based on observational data, might have a causal association with breast cancer-specific survival. We also aimed to investigate whether we could observe—or refute—an effect for the risk factors for which the association is not clear. We therefore performed a two-sample MR analysis using genetic variants and risk factor association summary estimates from the GWAS Catalog [22] and breast cancer survival summary estimates from the Breast Cancer Association Consortium (BCAC) cohort [23].

Methods

Selection of risk factors

We first considered the full list of breast cancer risk factors provided on the Cancer Research UK site [33] as of January 2020 (Additional file 1: Table S1). From this list of 25 factors, we identified nine factors for which genome-wide association study (GWAS) data were

available. Only GWASs that could be directly downloaded from GWAS Catalog [22] into TwoSampleMR [34] R package were considered. If there were multiple GWAS for one risk factor, we selected the study with the largest sample size from those that were predominantly of European ancestry (Table 1). We considered only genome-wide significant variants ($P < 5 \times 10^{-8}$) to ensure that the association with the risk factor was robust (first MR assumption). Only single-nucleotide polymorphisms (SNPs) were considered as the reference panel did not include other types of variants. Variants correlated with the most significant SNPs were removed so that only uncorrelated variants remained in the analysis ($r^2 < 0.001$). We calculated a priori power to detect an association at a significant level of 0.05 for each risk factor using the tool (<https://sb452.shinyapps.io/power>) [35]. We used the number of events ($n = 7054$) as sample size.

Breast cancer survival and genetic data

The breast cancer survival data was obtained from the Breast Cancer Association Consortium (BCAC). We analysed clinic-pathological data (database version 12) and genotype data from the OncoArray [36] and iCOGS arrays [37]. The analysis included 86,627 female patients of European ancestry diagnosed at age > 18 years with invasive breast cancer of any stage. The dataset included 7054 breast cancer-specific deaths. A total of 59,378 patients (4246 deaths) had ER-positive disease, and 13,692

(1733 deaths) had ER-negative disease. Genotypes for variants not present on the arrays were imputed using the Haplotype Reference Consortium [38] as reference panel. Details about the genotyping, sample quality control, and imputation procedure have been described previously [36, 39]. Our analyses were based on SNPs that were imputed with imputation $r^2 > 0.7$ and had minor allele frequency > 0.01 in at least one of the two datasets (iCOGS or OncoArray).

Breast cancer survival estimates

We took the SNPs referred to in Table 1 as genetic instruments for each of the nine risk factors. For every SNP, we performed survival analyses to obtain survival estimates as described previously [23]. The analyses included the full OncoArray and iCOGS datasets. Time at risk was calculated from the date of diagnosis with left truncation for prevalent cases. Follow-up was right censored on the date of death, last date known alive if death did not occur, or at 15 years after diagnosis, whichever came first [39]. We estimated the association between the genetic instruments and breast cancer-specific survival using Cox proportional hazards regression [40]. The models were stratified by study and included the first two ancestry informative principal components, based on the genotyping array data as previously described, to adjust for population structure [36, 37]. We analysed the OncoArray and iCOGS datasets separately and then combined the estimates using fixed-effect

Table 1 Description of the nine risk factors with available genetic data from GWAS

| Risk factor (units) | Number* of SNPs | GWAS Catalog accession number | Study reference | Sample size | Population |
|--|-----------------|-------------------------------|-----------------------------|---|---------------|
| Alcohol consumption (drinks per week) [†] | 98 | GCST007472 | Liu et al. [24] | 1,039,210 | 100% European |
| Body mass index (adult, kg/m ²) | 100 | GCST006368 | Hoffmann et al. [25] | 334,487 | 81% European |
| Height (adult, m) | 112 | GCST002647 | Wood et al. [26] | 253,288 | 100% European |
| Mammographic density (dense vs non-dense area) | 5 | GCST002667 | Lindström et al. [27] | 16,015** | 100% European |
| Menarche (age at onset) | 82 | GCST002541 | Perry et al. [28] | 182,413** | 100% European |
| Menopause (age at onset) | 34 | GCST005312 | Day et al. [29] | 69,626** | 100% European |
| Physical activity (overall physical activity time) | 3 | GCST006912 | Doherty et al. [30] | 91,105 | 100% European |
| Smoking behaviour (ever vs never) [‡] | 119 | GCST007327 | Karlsson Linnér et al. [31] | 518,633 | 100% European |
| Type 2 diabetes mellitus (yes vs no) | 95 | GCST006867 | Xue et al. [32] | 659,316 (62,892 cases and 596,424 controls) | 99% European |

*The number of SNPs here may be lower than in the GWAS due to filtering (see the “Methods” section)

**Partial overlap with samples from BCAC

[†]Defined as the average number of drinks a participant reported drinking each week, aggregated across all types of alcohol

[‡]This is a binary phenotype. Participants who reported ever being a regular smoker in their life (current or former) were coded “2”, while participants who reported never being a regular smoker in their life were coded “1”

meta-analyses [39]. Analyses were carried out for all invasive breast cancer and for estrogen receptor (ER)-positive and ER-negative disease separately. Additional file 2: Tables S1-S9 provides the full list of SNPs used and the corresponding estimates for the per-allele risk factor effect sizes and the per-allele survival log (hazard ratios).

MR statistical analyses and sensitivity diagnostics

We used the TwoSampleMR [34] R package to perform the two-sample MR analyses. We obtained the genetic instruments for the risk factors (MR-Base NHGRI-EBI GWAS Catalog [22], 29 August 2019 update), harmonised the SNP effects so they corresponded to the same allele for the risk factor and survival associations, and performed the sensitivity tests. We estimated the causal relationships between each of the sets of SNPs for the nine risk factors and breast cancer-specific survival using the inverse-variance weighted (IVW) method. We performed the analyses for all invasive breast cancer, ER-positive, and ER-negative separately. The association of BMI with breast cancer-specific survival was previously evaluated in an earlier, smaller version of the BCAC dataset ($n = 36,210$) [41]. In this analysis, we included more patients, updated follow-up, and a larger BMI GWAS genetic instrument. It has been suggested that the potential negative effect of BMI on survival is especially relevant in postmenopausal women [12]. Therefore, we also tested whether the BMI associations differed between pre- (age at diagnosis under 50 years, $n = 27,009$ with 2680 breast cancer-specific deaths) and postmenopausal women (age at diagnosis 50 years or older, $n = 59,617$ with 4374 breast cancer-specific deaths). Inclusion of even a small percentage of a different ethnic group can affect the interpretation and validity of the causal estimates [42]. Because the genetic instrument that we used for BMI had 19% of non-European participants, we performed an additional analysis using the BMI European-specific summary estimates from the same GWAS available at the author's supplementary material [25] (61 SNPs after filtering, Additional file 2: Table S10).

IVW assumes that none of the variants exhibit horizontal pleiotropy, which may not be true in practice. Therefore, we also used the MR-Egger regression method that allows variants to demonstrate unbalanced pleiotropic associations. That is, MR-Egger regression relaxes the requirement of no horizontal pleiotropy provided that the pleiotropic effects are not proportional to the effects of the variants on the risk factors of interest [8, 9]. In comparison to the IWM, the MR-Egger method's intercept is not constrained to zero and provides a statistical test of the extent to which this intercept differs from zero as a measure of unbalanced pleiotropic effects.

For the risk factors with a significant association based on the IVW method (false discovery rate [FDR] < 0.05), we ran the following sensitivity analyses: heterogeneity tests, funnel plots, and leave-one-out tests. To assess the robustness of the results of the IVW method, we applied other MR methods (simple mode, weighted median, and weighted mode). We also tested all associations by performing the analysis using a multivariable model. In the multivariable model, we used imputed phenotypes [43] and adjusted for the following known prognostic factors: age of the patients at diagnosis; tumour size; node status; distant metastasis status; grade; ER-, progesterone receptor, and HER2-status; and (neo) adjuvant chemotherapy, adjuvant anti-hormone therapy, and adjuvant trastuzumab. Because breast cancer survival can differ on the short or longer term, we also assessed whether or not the associations would hold for the 5-year horizon, which is typically used in breast cancer prognostication [44]. For this analysis, we reduced the follow-up time from 15 to 10 years ($n = 85,470$ with 6147 breast cancer-specific deaths) and 5 years ($n = 79,183$ with 3573 breast cancer-specific deaths). Both in the multivariable model and the shorter follow-up analyses, we performed the MR analyses separately for OncoArray and iCOGS datasets and meta-analysed the results.

Relationships between BMI, T2DM, and breast cancer survival

To ensure that the effects of BMI and T2DM were independent, we identified SNPs that overlapped between the genetic instruments for these risk factors. Two SNPs, rs7144011 and rs7903146, were present in both the BMI and T2DM instrumental variables, and 12 (six pairs) SNPs were in linkage disequilibrium (LD): rs2972144, rs4072096, rs1801282, rs1899951, rs2112347, rs2307111, rs4715210, rs72892910, rs244415, rs889398, rs6059662, and rs6142096. We removed those 14 SNPs from the analyses to reduce the likelihood of horizontal pleiotropy. To further isolate the association of T2DM alone, we performed a multivariable MR model [45] by additionally including the genetically predicted BMI score as a covariate in the analyses of T2DM.

Results

We found a significant association between genetic liability to T2DM and breast cancer-specific survival ($P < 0.05$, Table 2). For all breast cancers, T2DM was associated with worse breast cancer-specific survival (hazard ratio [HR] = 1.10, 95% confidence interval [CI] = 1.04–1.18, P value [P] = 0.003, FDR = 0.023) (Fig. 1 and Table 2). T2DM was also associated with worse breast cancer-specific survival when restricting to ER-positive cases. The effect in the ER-positive subtype was consistent (HR = 1.09, CI = 1.01–1.18, $P = 0.036$, FDR = 0.324)

Table 2 Effect of nine breast cancer risk factors on breast cancer-specific survival for all breast cancers, estrogen receptor (ER)-positive and ER-negative breast cancers. HR hazard ratio, CI 95% confidence interval, FDR false discovery rate

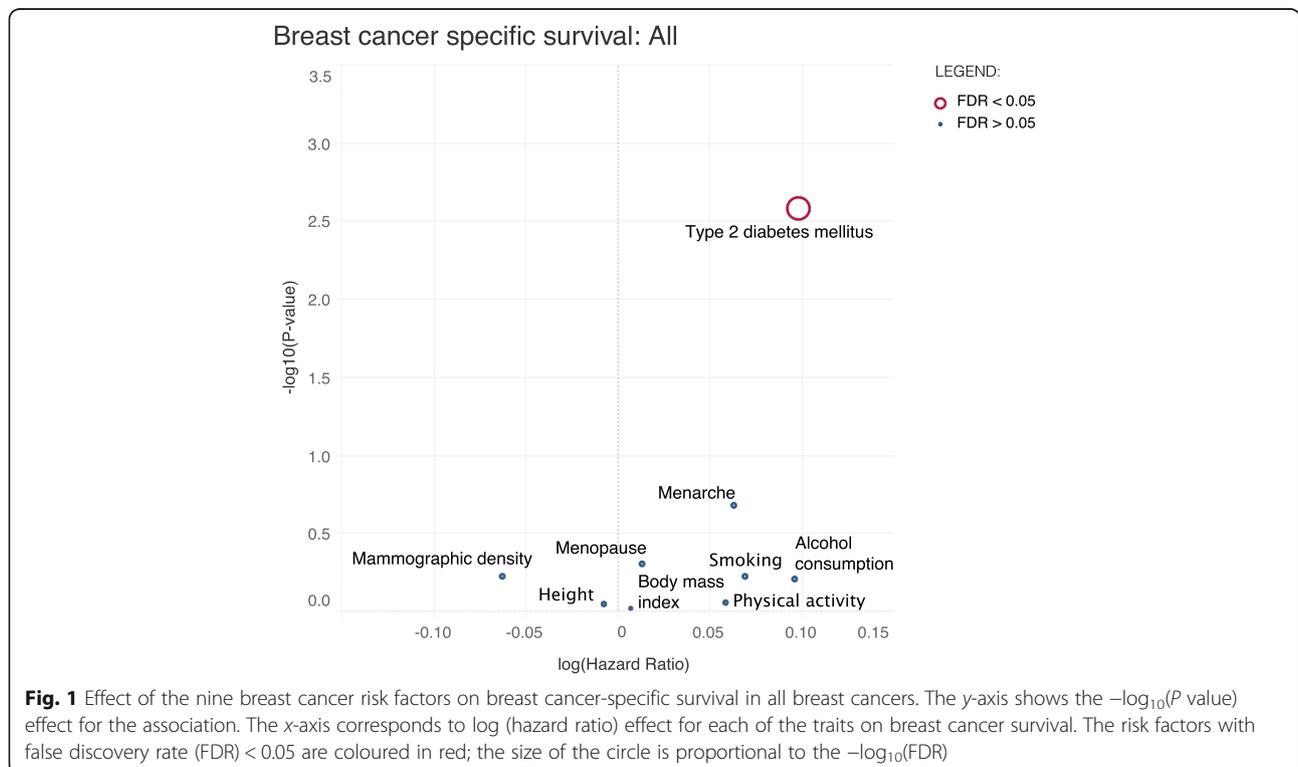
| Risk factor | All | | | | ER-positive | | | | ER-negative | | | |
|--|------|-----------|---------|------|-------------|-----------|---------|------|-------------|-----------|---------|------|
| | HR | CI | P value | FDR | HR | CI | P value | FDR | HR | CI | P value | FDR |
| Alcohol consumption (drinks per week) | 1.10 | 0.76–1.60 | 0.62 | 0.92 | 1.12 | 0.69–1.81 | 0.65 | 0.65 | 0.91 | 0.42–1.97 | 0.82 | 0.90 |
| Body mass index (adult, kg/m ²) | 1.01 | 0.81–1.25 | 0.95 | 0.95 | 1.10 | 0.83–1.45 | 0.52 | 0.62 | 1.14 | 0.72–1.82 | 0.57 | 0.88 |
| Height (adult, m) | 0.99 | 0.89–1.11 | 0.89 | 0.95 | 1.04 | 0.91–1.21 | 0.55 | 0.62 | 1.08 | 0.82–1.27 | 0.88 | 0.90 |
| Mammographic density (dense vs non-dense area) | 0.94 | 0.75–1.18 | 0.59 | 0.92 | 0.90 | 0.75–1.08 | 0.24 | 0.59 | 1.12 | 0.79–1.60 | 0.52 | 0.88 |
| Menarche (age at onset) | 1.07 | 0.97–1.18 | 0.21 | 0.92 | 1.06 | 0.94–1.21 | 0.34 | 0.61 | 1.06 | 0.86–1.29 | 0.59 | 0.88 |
| Menopause (age at onset) | 1.01 | 0.98–1.05 | 0.49 | 0.92 | 1.03 | 0.98–1.08 | 0.21 | 0.59 | 1.01 | 0.93–1.08 | 0.90 | 0.90 |
| Physical activity (overall physical activity time) | 1.06 | 0.52–2.16 | 0.87 | 0.95 | 1.66 | 0.48–5.71 | 0.43 | 0.62 | 0.36 | 0.09–1.54 | 0.17 | 0.77 |
| Smoking behaviour (ever vs never) | 1.07 | 0.83–1.38 | 0.59 | 0.92 | 0.82 | 0.58–1.16 | 0.26 | 0.59 | 1.53 | 0.92–2.54 | 0.10 | 0.77 |
| Type 2 diabetes mellitus (yes vs no) | 1.10 | 1.04–1.18 | 0.01 | 0.02 | 1.09 | 1.01–1.18 | 0.04 | 0.32 | 1.09 | 0.97–1.24 | 0.15 | 0.88 |

with the effect in all breast cancers. We did not observe associations at FDR < 0.05 (Table 2) between survival, for all breast cancer or by ER-subtype, and any of the other risk factors: alcohol consumption, BMI, height, mammographic density, menarche, menopause, physical activity, and smoking. The estimates we obtained from the models adjusted by other known prognostic factors (Additional file 1: Table S2) were comparable to the initial unadjusted analyses for all risk factors. Under the current sample size of our study ($n = 86,627$ and 7054 events), the power to detect a causal association varied considerably between risk factors (Additional file 1:

Table S3). The estimated power was the largest for age at menopause and lowest for physical activity.

Genetic association between BMI by menopausal status and breast cancer-specific survival

We found no association between BMI and breast cancer-specific survival in any of the analysed subtypes, nor by menopausal status ($P > 0.05$): premenopausal (HR = 1.06, CI = 0.78–1.44, $P = 0.710$) or postmenopausal women (HR = 1.02, CI = 0.80–1.30, $P = 0.899$). The estimate using the European-specific BMI genetic



instrument (HR = 1.14, CI = 0.94–1.38, $P = 0.174$) was also not significant.

Genetic association between T2DM and breast cancer-specific survival

The HR estimate for T2DM and survival among all invasive breast cancers (HR = 1.10) was higher than that for either ER-subtype individually (ER-positive: HR = 1.09; ER-negative: HR = 1.09). This reflected the fact that the patients without ER-status information ($n = 13,557$) had a larger risk estimate (HR = 1.19, CI = 1.02–1.39, $P = 0.023$).

To further validate the association between T2DM and breast cancer-specific survival, we performed the analysis using a shorter follow-up. The results were significant and similar to the main analysis both for 10-year (HR = 1.12, CI = 1.05–1.19, $P = 0.0006$) and for 5-year follow-up (HR = 1.13, CI = 1.04–1.23, $P = 0.005$). We also tested the association in a model adjusted by other known prognostic factors. The association of T2DM with breast cancer-specific survival in the adjusted model was still significant (HR = 1.10, CI = 1.02–1.18, $P = 0.013$), and the effect size remained similar to the main T2DM analysis (HR = 1.10, CI = 1.04–1.18, $P = 0.003$). Finally, we tried to replicate the result using another large and well-powered GWAS, i.e. the T2DM summary estimates from the DIAGRAM GWAS

which is a large meta-analysis of 32 studies comprising data for 898,130 individuals (74,124 T2DM cases and 824,006 controls) of European ancestry [46]. The genetic instrument for this dataset included 152 SNPs (12 SNPs overlapping with the T2DM genetic instrument we initially used, Additional file 2: Table S11). The association of T2DM with breast cancer-specific survival using the replication dataset was significant (HR = 1.18, CI = 1.04–1.33, $P = 0.009$) and similar to the initial result (HR = 1.10).

Association between T2DM and breast cancer-specific survival with BMI adjustment

To explore the potential confounding effect of BMI with T2DM, we performed an analysis adjusting for genetically predicted BMI. The effect of BMI in this analysis was not significant (HR = 1.02, CI = 0.85–1.24, $P = 0.809$), and the effect of T2DM on survival was similar (HR = 1.10, CI = 1.04–1.17, $P = 0.002$) to the main T2DM analysis (HR = 1.10, CI = 1.04–1.18, $P = 0.003$).

Causal association between T2DM and breast cancer-specific survival

We used different variations of the MR method to assess possible violations of the MR assumptions. Figure 2 shows that the range of MR methods used (simple

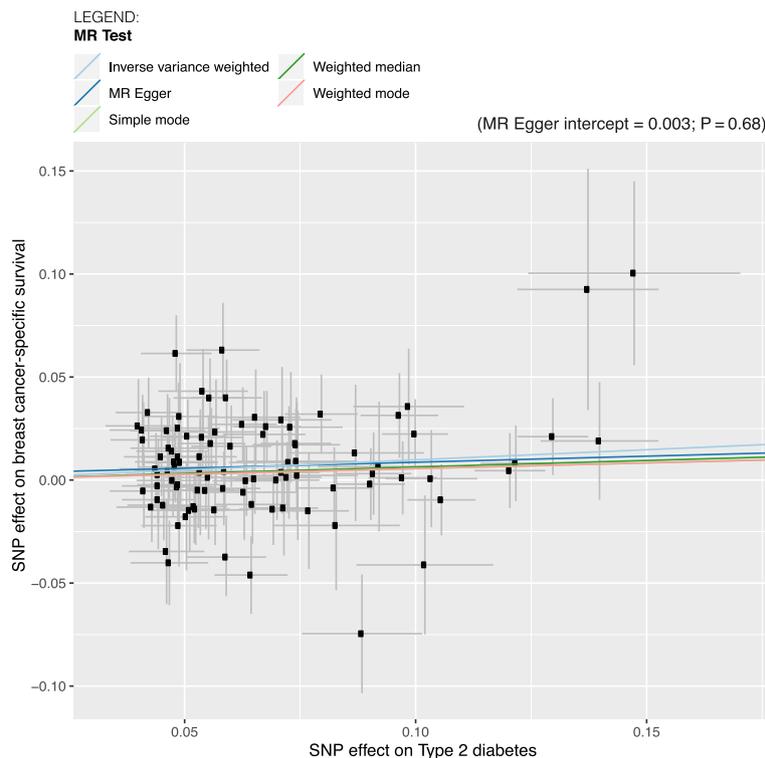


Fig. 2 Plot showing the effect sizes of the SNP effects on breast cancer-specific survival for all breast cancers (y-axes) and the SNP effects on T2DM (x-axes) with 95% confidence intervals. Each dot represents one of the 95 SNPs used in the T2DM genetic instrument. The slopes indicate the estimate for each of the five different MR tests

mode, weighted median, and weighted mode) to assess the sensitivity of the findings all gave similar effect size estimates. Additionally, there was no evidence of pleiotropy based on the MR-Egger intercept test (MR-Egger intercept = 0.003, $P = 0.68$, Fig. 2). In analyses using funnel plot (Additional file 1: Figure S1) and a leave-one-out test (Additional file 1: Figure S2), there was no indication for violation of the assumptions, nor that the association was driven by any particular SNP.

Discussion

We performed a Mendelian randomisation analysis to explore the potential causal effects on breast cancer-specific survival of nine established risk factors for breast cancer: alcohol consumption, BMI, height, mammographic density, menarche, menopause, physical activity, smoking, and T2DM. We used survival estimates from 86,627 European breast cancer patients with invasive breast cancer (by far the largest such dataset) and summary data from the GWAS Catalog for the nine risk factors. We used the IVW method to estimate causal effects and performed a wide range of sensitivity analyses to test the robustness of our findings.

Our analysis showed an association between genetic liability to T2DM and worse breast cancer-specific survival. The IVW method result was consistent with the results of other complementary MR-methods, and the performed sensitivity analyses did not give any statistical indication for violations of the MR assumptions. Additionally, the T2DM GWAS used was reasonably powered, with an estimated heritability of $\sim 20\%$ [32], supporting the relevance assumption. There was no evidence that the SNPs were associated with breast cancer survival (exclusion restriction). Finally, the association remained significant when adjusting for other known prognostic factors and when shortening the follow-up time to 10 and 5 years.

Because obesity and T2DM share some biological features such as elevated insulin levels, hypertension, and chronic inflammation [47] and since higher BMI has been associated with increased incidence of T2DM [48], we explored a possible interaction between the two risk factors. First, we ensured that there were no common SNPs between the T2DM and BMI genetic instruments or SNPs in LD that could be driving the association. Second, we performed BMI-adjusted analyses which also showed that the association was being driven by T2DM and not by BMI.

Earlier literature suggests an association between diabetes and worse breast cancer-specific survival [49–51]. There is no clear evidence linking diabetes to any particular ER-status specific breast cancer subtype [52] that could explain the poorer survival in women with T2DM. The increased mortality in patients with T2DM might

be explained by the effect of insulin resistance or hyperinsulinemia, since breast cancer cells might have a selective growth advantage because of insulin receptor overexpression [53, 54]. However, to our knowledge, no functional studies to evaluate this have yet been carried out. An important point to consider when interpreting the results is that, when using a binary risk factors such as T2DM, the genetic instrument estimate will only represent the average causal effect of the exposure in a fraction of the studied population (named “genetic compliers”). Additionally, the latter would only be true assuming that the monocity assumption is plausible, which means that increasing number of alleles for an individual would increase (or maintain constant) the risk of having T2DM [55].

All the other risk factors gave null results. Some of these may reflect the fact that there is no true association, but others may be underpowered since the fraction of variation of the risk factor explained by the genetic instrument was too small. The heritability explained by identified SNPs, and hence the power of the genetic instruments, varies substantially between risk factors, e.g. $\sim 20\%$ for T2DM [32] versus only 1% for the mammographic density GWAS [27]. In addition, we only kept genome-wide significant SNPs and dropped all SNPs in LD or with low imputation quality in the BCAC dataset, so the explained variation that we could utilise was smaller. As GWAS become larger and more powerful genetic instruments are available, it may be possible to find associations that could not be identified here. However, for those risk factors with a predicted small genetic component (e.g. physical activity), their association with breast cancer survival might not be assessable using an MR framework [8]. A potential limitation of our study is that some patients in the breast cancer survival dataset were also included in the GWASs for the risk factors, mammographic density ($\sim 2.5\%$ overlap) and age at menarche ($\sim 27\%$) and menopause ($\sim 21\%$). However, because the genetic instruments of age at menarche and menopause were relatively strong and there was little overlap for mammographic density, we may expect the bias caused by patient overlap to be small [56]. Finally, another potential reason for which we did not observe association for some risk factors might be due to selection bias. This type of collider bias can lead to an under- or overidentification of genetic risk factors for breast cancer survival due to a relationship between the genetic risk factor concerned and breast cancer incidence [57]. This could be the case for BMI, age at menopause and menarche, or height, which have been causally associated with breast cancer risk [58]. For other risk factors such as T2DM or smoking, MR studies of incidence could not provide evidence for a

causal association [59, 60], which makes these genetic instruments less likely to be affected by selection bias.

To further explore the link between BMI and breast cancer survival, we also tested separately for pre- and postmenopausal status, but there was no indication for an association in any of the menopausal groups. Despite the evidence for an association between BMI and breast cancer survival from observational studies [12, 13], our analysis on BMI and breast cancer-specific survival did not confirm this. A possible explanation is that obesity is associated with other comorbid conditions [48] that lead to poorer overall, but no breast cancer-specific survival. Additionally, it has been suggested that obese patients might receive suboptimal chemotherapy treatment compared to regular weight women [61] and tumours are usually detected at a later stage in obese patients [62]. This would, if insufficiently corrected for, lead to an association between high BMI and worse breast cancer-specific survival in observational, but not in MR, studies. The different observations of the relationship between BMI and survival from MR versus observational studies resemble those of genetic BMI and breast cancer risk [63], which were also deviant from epidemiological studies. To date, there is not a clear answer as to whether and how high BMI directly influences the biology of cancer [64].

From a clinical point of view, our analysis suggests that genetic liability to T2DM may contribute to variation in breast cancer outcomes in women of European ancestry. Such a genetic predictor might be included in prognostication models aimed at identifying women most likely to benefit from specific interventions. Furthermore, even though T2DM has a genetic component, it is also influenced by environmental and lifestyle factors and is potentially preventable [65]. Although our study does not address this directly, it seems sensible to recommend intensified management of T2DM, including lifestyle changes, in breast cancer patients.

The main strength of our study is the use of the biggest breast cancer dataset available so far and the use of SNPs as genetic instruments to reduce potential confounding. Despite including more than 7000 breast cancer-specific deaths in the analyses, our study was not well powered especially for the analysis within the subset of ER-negative tumours (as indicated by the broad confidence intervals). Additional findings might be possible when there are larger sample sizes available and a more complete follow-up. We also lacked power to detect associations for certain risk factors that had only a handful of SNPs in their genetic instruments such as mammographic density and physical activity. Finally, our results are applicable to women of European ancestry only. In order to be able to generalise these findings to other

ancestry groups, larger breast cancer datasets are needed for the other ethnicities.

Conclusion

This two-sample MR analysis suggests that genetic liability to T2DM might be a cause of reduced breast cancer-specific survival. Our study provides further evidence for the importance of promoting a healthier lifestyle to improve survival in breast cancer patients.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12916-020-01797-2>.

Additional file 1: Table S1. List of breast cancer risk factors as indicated by Cancer Research UK. Information in this table was taken directly from: <https://www.cancerresearchuk.org/about-cancer/breast-cancer/risks-causes/risk-factors> (January 2020). **Table S2.** Comparison of the effect of nine breast cancer risk factors on breast cancer-specific survival for all breast cancers in the unadjusted model (left) and in the adjusted model (right). The model was adjusted for the known prognostic factors: age of the patients at diagnosis, tumour size, node status, distant metastasis status, grade, ER-, progesterone receptor and HER2-status and (neo) adjuvant chemotherapy, adjuvant anti-hormone therapy and adjuvant trastuzumab. HR = Hazard Ratio. CI = 95% Confidence Interval.

Table S3. Power (%) estimation by a range of Hazard Ratios (HR) for the analysis of MR associations between nine breast cancer risk factors and breast cancer-specific survival in all breast cancers. **Figure S1.** Funnel plot for T2DM and breast cancer-specific survival. The plot shows the effect estimate (b) of a particular SNP against the SNP expected precision (1/Standard Error (SE)). Asymmetry in the funnel plot is an indication of horizontal pleiotropy. The dark and light blue lines represent the MR-Egger and Inverse variance weighted slopes respectively. **Figure S2.** Leave-one-out plot for T2DM and breast cancer specific-survival showing the estimate effect by sequentially dropping one SNP at a time. Each black dot in the forest plot represents the MR results (IVW method) excluding that particular SNP. The result including all SNPs is shown in red at the bottom of the plot.

Additional file 2: SNPs used in the analyses for the nine risk factors. The risk factor estimates (beta and standard error (SE)) and breast cancer-specific survival estimates for each SNP are included. **Table S1.** Alcohol consumption. **Table S2.** Body mass index. **Table S3.** Height. **Table S4.** Mammographic density. **Table S5.** Menarche. **Table S6.** Menopause. **Table S7.** Physical activity. **Table S8.** Smoking behaviour. **Table S9.** Type 2 diabetes mellitus. **Table S10.** Body mass index European-specific. **Table S11.** Type 2 diabetes mellitus replicate.

Abbreviations

BMI: Body mass index; BCAC: Breast Cancer Association Consortium; CI: Confidence interval; ER: Estrogen receptor; FDR: False discovery rate; GWAS: Genome-wide association study; HR: Hazard ratio; IWV: Inverse-variance weighted; LD: Linkage disequilibrium; MR: Mendelian randomisation; *P*: *P* value; RCT: Randomised control trial; SNP: Single-nucleotide polymorphism; T2DM: Type 2 diabetes mellitus

Acknowledgements

BCAC: We thank all the individuals who took part in these studies and all the researchers, clinicians, technicians, and administrative staff who have enabled this work to be carried out. We acknowledge all contributors to the COGS and OncoArray study design, chip design, genotyping, and genotype analyses.

Authors' contributions

M.K.S., S.C., and M.E.G. designed the study. M.E.G. performed the main data analyses and drafted the initial manuscript. A.M. performed the adjusted analysis. M.K.S., S.C., and M.E.G. interpreted the data. All authors were

involved in the data collection, commented on the drafts, and approved the final version of the manuscript.

Funding

BCAC is funded by Cancer Research UK (C1287/A16563, C1287/A10118), by the European Union's Horizon 2020 Research and Innovation Programme (grant numbers 634935 and 633784 for BRIDGES and B-CAST, respectively), and by the European Community's Seventh Framework Programme under grant agreement number 223175 (grant number HEALTH-F2-2009-223175) (COGS). The EU Horizon 2020 Research and Innovation Programme funding source had no role in the study design, data collection, data analysis, data interpretation, or writing of the report. Genotyping of the OncoArray was funded by the NIH Grant U19 CA148065, and Cancer UK Grant C1287/A16563 and the PERSPECTIVE project supported by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research (grant GPH-129344), and the Ministère de l'Économie, Science et Innovation du Québec through Genome Québec and the PSRSIIRI-701 grant, and the Quebec Breast Cancer Foundation. Funding for the iCOGS infrastructure came from the European Community's Seventh Framework Programme under grant agreement no. 223175 (HEALTH-F2-2009-223175) (COGS), Cancer Research UK (C1287/A10118, C1287/A10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692, C8197/A16565), the National Institutes of Health (CA128978) and Post-Cancer GWAS initiative (1U19 CA148537, 1U19 CA148065 and 1U19 CA148112 - the GAME-ON initiative), the Department of Defence (W81XWH-10-1-0341), the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer, and Komen Foundation for the Cure, the Breast Cancer Research Foundation, and the Ovarian Cancer Research Fund. M.E.G was funded by the Dutch Cancer Society (grant 2015-7632).

Availability of data and materials

Not applicable.

Ethics approval and consent to participate

The study was performed in accordance with the Declaration of Helsinki. Summary estimates were used from previously reported studies that followed the appropriate institutional review and patient consent procedures and followed the procedure the Data Access Coordination Committee (DACC) of BCAC (<http://bcac.ccg.medschl.cam.ac.uk/>).

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Division of Molecular Pathology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands. ²Division of Molecular Carcinogenesis, The Netherlands Cancer Institute - Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands. ³Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany. ⁴University Medical Center Hamburg-Eppendorf, University Cancer Center Hamburg (UCCH), Cancer Epidemiology Group, Hamburg, Germany. ⁵MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK. ⁶Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK. ⁷Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, USA. ⁸Copenhagen University Hospital, Copenhagen General Population Study, Herlev and Gentofte Hospital, Herlev, Denmark. ⁹Copenhagen University Hospital, Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Herlev, Denmark. ¹⁰Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. ¹¹Department of Oncology, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, UK. ¹²Department of Public Health and Primary Care, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, UK. ¹³Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands.

Received: 15 May 2020 Accepted: 25 September 2020

Published online: 17 November 2020

References

1. Polyak K. Heterogeneity in breast cancer. *J Clin Invest*. 2011;121:3786–8.
2. Phung MT, Tin Tin S, Elwood JM. Prognostic models for breast cancer: a systematic review. *BMC Cancer*. 2019;19:230.
3. Hellmann SS, Thygesen LC, Tolstrup JS, Grønbaek M. Modifiable risk factors and survival in women diagnosed with primary breast cancer: results from a prospective cohort study. *Eur J Cancer Prev*. 2010;19:366–73.
4. Barnett GC, Shah M, Redman K, Easton DF, Ponder BAJ, Pharoah PDP. Risk factors for the incidence of breast cancer: do they affect survival from the disease? *J Clin Oncol*. 2008;26:3310–6.
5. Reeves GK, Patterson J, Vessey MP, Yeates D, Jones L. Hormonal and other factors in relation to survival among breast cancer patients. *Int J Cancer*. 2000;89:293–9.
6. Hayes SC, Steele ML, Spence RR, Gordon L, Battistutta D, Bashford J, et al. Exercise following breast cancer: exploratory survival analyses of two randomised, controlled trials. *Breast Cancer Res Treat*. 2018;167:505–14.
7. Stagl JM, Lechner SC, Carver CS, Bouchard LC, Gudenkauf LM, Jutagir DR, et al. A randomized controlled trial of cognitive-behavioral stress management in breast cancer: survival and recurrence at 11-year follow-up. *Breast Cancer Res Treat*. 2015;154:319–28.
8. Davies NM, Holmes MV, Davey SG. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ*. 2018;362.
9. Smith GD, Borges M-C, Bowden J, Evans DM, Haycock P, Hemani G, et al. Recent developments in Mendelian randomization studies. *Curr Epidemiol Reports*. 2017;4:330–45.
10. Passarelli MN, Newcomb PA, Hampton JM, Trentham-Dietz A, Titus LJ, Egan KM, et al. Cigarette smoking before and after breast cancer diagnosis: mortality from breast cancer and smoking-related diseases. *J Clin Oncol*. 2016;34:1315–22.
11. Duan W, Li S, Meng X, Sun Y, Jia C. Smoking and survival of breast cancer patients: a meta-analysis of cohort studies. *Breast*. 2017;33:117–24.
12. Protani M, Coory M, Martin JH. Effect of obesity on survival of women with breast cancer: systematic review and meta-analysis. *Breast Cancer Res Treat*. 2010;123:627–35.
13. Picon-Ruiz M, Morata-Tarifa C, Valle-Goffin JJ, Friedman ER, Slingerland JM. Obesity and adverse breast cancer risk and outcome: mechanistic insights and strategies for intervention. *CA Cancer J Clin*. 2017;67:378–97.
14. Mu L, Zhu N, Zhang J, Xing F, Li D, Wang X. Type 2 diabetes, insulin treatment and prognosis of breast cancer. *Diabetes Metab Res Rev*. 2017;33:e2823.
15. Lega IC, Austin PC, Fischer HD, Fung K, Krzyzanowska MK, Amir E, et al. The impact of diabetes on breast cancer treatments and outcomes: a population-based study. *Diabetes Care*. 2018;41:755–61.
16. Orgéas CC, Hall P, Rosenberg LU, Czene K. The influence of menstrual risk factors on tumor characteristics and survival in postmenopausal breast cancer. *Breast Cancer Res*. 2008;10:R107.
17. Olsson Å, Sartor H, Borgquist S, Zackrisson S, Manjer J. Breast density and mode of detection in relation to breast cancer specific survival: a cohort study. *BMC Cancer*. 2014;14:229.
18. Olsen AH, Bihmann K, Jensen M-B, Vejborg I, Lyng E. Breast density and outcome of mammography screening: a cohort study. *Br J Cancer*. 2009;100:1205–8.
19. van der Waal D, Verbeek ALM, Broeders MJM. Breast density and breast cancer-specific survival by detection mode. *BMC Cancer*. 2018;18:386.
20. Courneya KS, Segal RJ, McKenzie DC, Dong H, Gelmon K, Friedenreich CM, et al. Effects of exercise during adjuvant chemotherapy on breast cancer outcomes. *Med Sci Sports Exerc*. 2014;46:1744–51.
21. Ali AMG, Schmidt MK, Bolla MK, Wang Q, Gago-Dominguez M, Castela J, et al. Alcohol consumption and survival after a breast cancer diagnosis: a literature-based meta-analysis and collaborative analysis of data for 29,239 cases. *Cancer Epidemiol Biomark Prev*. 2014;23:934–45.
22. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res*. 2019;47:D1005–12.
23. Escala-Garcia M, Abraham J, Andrusis IL, Anton-Culver H, Arndt V, Ashworth A, et al. A network analysis to identify mediators of germline-driven differences in breast cancer prognosis. *Nat Commun*. 2020;11:312.

24. Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet.* 2019;51:237–44.
25. Hoffmann TJ, Choquet H, Yin J, Banda Y, Kvale MN, Glymour M, et al. A large multiethnic genome-wide association study of adult body mass index identifies novel loci. *Genetics.* 2018;210:499–515.
26. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet.* 2014;46:1173–86.
27. Lindström S, Thompson DJ, Paterson AD, Li J, Gierach GL, Scott C, et al. Genome-wide association study identifies multiple loci associated with both mammographic density and breast cancer risk. *Nat Commun.* 2014;5:5303.
28. Perry JR, Day F, Elks CE, Sulem P, Thompson DJ, Ferreira T, et al. Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature.* 2014;514:92–7.
29. Day FR, Ruth KS, Thompson DJ, Lunetta KL, Pervjakova N, Chasman DI, et al. Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. *Nat Genet.* 2015;47:1294–303.
30. Doherty A, Smith-Byrne K, Ferreira T, Holmes MV, Holmes C, Pulit SL, et al. GWAS identifies 14 loci for device-measured physical activity and sleep duration. *Nat Commun.* 2018;9:5257.
31. Karlsson Linnér R, Biroli P, Kong E, Meddens SFW, Wedow R, Fontana MA, et al. Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nat Genet.* 2019;51:245–57.
32. Xue A, Wu Y, Zhu Z, Zhang F, Kemper KE, Zheng Z, et al. Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nat Commun.* 2018;9:2941.
33. Cancer Research UK. Risk factors for breast cancer. <https://www.cancerresearchuk.org/about-cancer/breast-cancer/risks-causes/risk-factors>.
34. 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;7:1–29.
35. Burgess S. Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome. *Int J Epidemiol.* 2014;43:922–9.
36. Amos CI, Dennis J, Wang Z, Byun J, Schumacher FR, Gayther SA, et al. The OncoArray consortium: a network for understanding the genetic architecture of common cancers. *Cancer Epidemiol Biomark Prev.* 2017;26:126–35.
37. 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;551:92–4.
38. Frye FL, Cucuel JP. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet.* 2016;48:1279–83.
39. Escala-Garcia M, Guo Q, Dörk T, Canisius S, Keeman R, Dennis J, et al. Genome-wide association study of germline variants and breast cancer-specific mortality. *Br J Cancer.* 2019;120:647–57.
40. Cox DR, Hinkley DV. *Theoretical Statistics.* Boston: Springer US; 1974.
41. Guo Q, Burgess S, Turman C, Bolla MK, Wang Q, Lush M, et al. Body mass index and breast cancer survival: a Mendelian randomization analysis. *Int J Epidemiol.* 2017;46:1891–902.
42. Burgess S, Davey Smith G, Davies NM, Dudbridge F, Gill D, Glymour MM, et al. Guidelines for performing Mendelian randomization investigations. *Wellcome Open Res.* 2020;4:186.
43. van Buuren S. Flexible imputation of missing data, second edition. Chapman and Hall/CRC: Second edition. Boca Raton: CRC Press. 2019. 2018.
44. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. *Lancet.* 2011;378:771–84.
45. Burgess S, Thompson SG. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *Am J Epidemiol.* 2015;181:251–60.
46. Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet.* 2018;50:1505–13.
47. Jiralerspong S, Kim ES, Dong W, Feng L, Hortobagyi GN, Giordano SH. Obesity, diabetes, and survival outcomes in a large cohort of early-stage breast cancer patients. *Ann Oncol.* 2013;24:2506–14.
48. Guh DP, Zhang W, Bansback N, Amarsi Z, Birmingham CL, Anis AH. The incidence of co-morbidities related to obesity and overweight: a systematic review and meta-analysis. *BMC Public Health.* 2009;9:88.
49. Goodwin PJ, Ennis M, Pritchard KI, Trudeau ME, Koo J, Madarnas Y, et al. Fasting insulin and outcome in early-stage breast cancer: results of a prospective cohort study. *J Clin Oncol.* 2002;20:42–51.
50. Lipscombe LL, Goodwin PJ, Zinman B, McLaughlin JR, Hux JE. The impact of diabetes on survival following breast cancer. *Breast Cancer Res Treat.* 2008;109:389–95.
51. Tsilidis KK, Kasimis JC, Lopez DS, Ntzani EE, Ioannidis JPA. Type 2 diabetes and cancer: umbrella review of meta-analyses of observational studies. *BMJ.* 2015;350:g7607.
52. Bronsveld HK, Jensen V, Vahl P, De Bruin ML, Cornelissen S, Sanders J, et al. Diabetes and breast cancer subtypes. *PLoS One.* 2017;12:e0170084.
53. Papa V, Belfiore A. Insulin receptors in breast cancer: biological and clinical role. *J Endocrinol Invest.* 1996;19:324–33.
54. Coughlin SS, Calle EE, Teras LR, Petrelli J, Thun MJ. Diabetes mellitus as a predictor of cancer mortality in a large cohort of US adults. *Am J Epidemiol.* 2004;159:1160–7.
55. Burgess S, Labrecque JA. Mendelian randomization with a binary exposure variable: interpretation and presentation of causal estimates. *Eur J Epidemiol.* 2018;33:947–52.
56. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. *Genet Epidemiol.* 2016;40:597–608.
57. Paternoster L, Tilling K, Davey SG. Genetic epidemiology and Mendelian randomization for informing disease therapeutics: conceptual and methodological challenges. *PLoS Genet.* 2017;13:e1006944.
58. Pierce BL, Kraft P, Zhang C. Mendelian randomization studies of cancer risk: a literature review. *Curr Epidemiol Reports.* 2018;5:184–96.
59. Yuan S, Kar S, Carter P, Vithayathil M, Mason AM, Burgess S, et al. Is type 2 diabetes causally associated with cancer risk? Evidence from a two-sample Mendelian randomization study. *Diabetes.* 2020;69:1588–96.
60. Larsson SC, Carter P, Kar S, Vithayathil M, Mason AM, Michaëlsson K, et al. Smoking, alcohol consumption, and cancer: a mendelian randomisation study in UK Biobank and international genetic consortia participants. *PLoS Med.* 2020;17:e1003178.
61. Griggs JJ, Sorbero MES, Lyman GH. Undertreatment of obese women receiving breast cancer chemotherapy. *Arch Intern Med.* 2005;165:1267.
62. Cui Y, Whiteman MK, Flaws JA, Langenberg P, Tkaczuk KH, Bush TL. Body mass and stage of breast cancer at diagnosis. *Int J Cancer.* 2002;98:279–83.
63. Shu X, Wu L, Khankari NK, Shu X-O, Wang TJ, Michailidou K, et al. Associations of obesity and circulating insulin and glucose with breast cancer risk: a Mendelian randomization analysis. *Int J Epidemiol.* 2019;48:795–806.
64. Azvolinsky A. Cancer prognosis: role of BMI and fat tissue. *JNCI J Natl Cancer Inst.* 2014;106.
65. Mokdad AH, Ford ES, Bowman BA, Dietz WH, Vinicor F, Bales VS, et al. Prevalence of obesity, diabetes, and obesity-related health risk factors, 2001. *JAMA.* 2003;289:76.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

