High oestradiol levels increase the risk of breast and endometrial cancer

Åsa Johansson[1], Daniel Schmitz[1], Julia Höglund[1], Fatemeh Hadizadeh[1], Torgny Karlsson[1], Weronica E. Ek[1*]

[1] Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Box 815, 75108, Uppsala. Sweden

***Corresponding author:** Weronica E. Ek: Weronica.ek@igp.uu.se, BMC, box815, 751 08 Uppsala, +46703519004

Running title: Oestradiol is causal for cancer

Keywords: Mendelian Randomization, Oestradiol, Breast cancer, Endometrial cancer, Ovarian cancer

Word count: 4396

Number of figures and tables: 4

Funding

This work was funded by the Swedish Research Council 2019-01497 (ÅJ), The Swedish Cancer Society 19 0383 Pj (ÅJ), the Marcus Borgström Foundation (WEE), K and O F Hedströms Foundation (WEE), the Åke Wiberg Foundation M19-0349/M20-0057 (WEE), A and M Rudbergs Foundation (WEE), and Vleugels Foundation (WEE).
The funding sources had no influence and took no part in the design or conduct of this research.

The authors declare no potential conflicts of interest

Abstract

Background: High levels of oestrogen are associated with increased risk of breast and endometrial cancer and has been suggested to also play a role in the development of ovarian cancer. Cancerogenic effects of oestradiol, the most prominent form of oestrogen, has been highlighted as a side effect of oestrogen only menopausal hormone therapy. However, whether high levels of endogenous oestrogens, produced within the body, promote cancer development, has not been fully established.

Methods: Here we performed Mendelian Randomization (MR) analyses to estimate the effect of endogenous oestrogen on the risk of developing breast, endometrial and ovarian cancer.

Results: We showed that high levels of oestradiol increase the risk of breast (OR=1.19 [95% CI 1.03-1.38]) and endometrial cancer (OR=1.45 [95% CI 1.14-1.83]) but we could not establish a link to the risk of ovarian cancer (OR=1.20 [95% CI 0.99-1.46]). We thereby suggest that individual variation in oestrogen production is indeed a causal determinant of cancer risk, which agrees with previous research.

Conclusion: This is the first MR study to identify an effect of oestradiol on breast and endometrial cancer using multiple instruments, with effects estimated in women.

Impact: These results can be of importance for future precision medicine interventions to prevent cancer in women.

Introduction

Oestrogens are hormones that are important for the development and control of the primary and secondary sexual characteristics in both genders, and play a key role in the tuning of the reproductive system in women [1,2]. Oestrogen, along with progesterone, broadly contribute to the function and development of a large number of human tissues including endometrium, breast, bone, brain and cardiovascular system [3]. Notably, oestrogen promotes growth in normal endometrial and breast luminal epithelial tissues [3,4]. Several studies have demonstrated that long-lasting exposure or higher levels of oestrogen, as a result of oestrogen medication, early menarche, late menopause, nulliparity, and obesity, is a shared origin of different malignancies including breast, ovarian and endometrial cancer [5–7].

The predominant source of oestrogen production in the premenopausal women are ovaries, while in the postmenopausal women, subcutaneous adipose tissue is the major origin [5]. In premenopausal women, oestrogen is produced by the growing ovarian follicle to stimulate endometrial thickness, so as to prepare it for the implantation of the fertilized egg. At the end of this phase and following a peak in the secretion of oestrogen, ovulation occurs and the corpus luteum will be the source of oestrogen production during the luteal phase. If no fertilization takes place, the corpus luteum becomes inactive and at the end of this phase the oestrogen levels drop and the menstruation starts. Progesterone is exclusively secreted during the luteal phase to oppose the oestrogen-induced growth of the endometrium [4,8,9].

In breast tissue, oestrogen plays an important role in development, proliferation and differentiation [10,11]. Oestradiol, the most potent oestrogen, is the biologically active type of oestrogen in both pre- and post-menopausal women [5]. Previous studies have shown an association between higher risk of breast cancer among women that have high blood levels of oestradiol both before [12,13] and after menopause [14,15]. Still, if high natural oestradiol levels in the body have a causal effect on breast cancer, or whether the high oestradiol levels are just a consequence of cancer progression, has not been clearly described.

Endometrial cancer is the most common gynaecological cancer and is a hormonally-responsive tumour [4]. Endometrial cancer risk increases with use of menopausal hormonal treatment (MHT) that includes oestrogen only. However, this risk can be reduced if the treatment is combined with (opposed by) progesterone [16,17]. Instead, a protective effect on

both endometrial and ovarian cancers have been identified in oral contraceptive users [18,19]. Oral contraceptives contain oestradiol in combination with a synthetic form of progesterone (progestin), or progestin only. The protective effect of oral contraceptives has been suggested to be due to fewer ovulations [20], since the peak in oestrogen secretion that triggers tissue growth in association with the ovulation is depleted during oral contraceptive use. Ovarian cancer is the most fatal of the female gynaecological cancers worldwide with no screening test and late-stage diagnosis [17 21]. Epidemiological studies have suggested a strong role for the oestrogen activity as well as the duration of exposure to it in the initiation, pathogenesis, and progression of ovarian cancer [17].

In addition, driving from experimental data, hormone-associated cancers such as breast, endometrial and ovarian cancer share several pathways for oestrogen regulation [17]. Even though oestrogen has been linked to all three cancer types, there is a lack of knowledge about whether the body's own production of oestrogen promotes the development of breast, endometrial and ovarian cancer. Establishing such causal relations is important in order to determine the pathogenetic mechanisms and to prevent cancers. One difficulty when studying risk factors, such as oestrogen levels on cancer, is to distinguish correlation from causation. Mendelian randomization (MR) is an instrumental variable approach that can be used to disentangle the effect that oestradiol exerts on the risk of developing cancer. In MR, germ-line genetic variants are used as instrumental variables. Thereby, the MR estimate is not affected by reversed causation, since genetic variants are assorted during formation of gametes prior to conception and are not confounded by lifestyle or environmental factors. MR is thereby an ideal approach for estimating causal effects. In MR, the instrumental variables must fulfil three fundamental assumptions in order to be valid: 1) the variant must be associated with the exposure (oestradiol), 2) the variant should not be associated with any potential confounder in the exposure – outcome relation, 3), and the variant should not be directly associated with the outcome (cancer) [22]. Two previous MR studies, both including only one genetic variant, close to the *CYP19A1* gene, that was identified to be associated with oestradiol levels primally in males, identified a causal effect of higher oestradiol levels on endometrial cancer risk [23,24]. MR studies based on a single genetic variant may be greatly biased, as a result of undetectable pleiotropic effects [25]. One previous study used additional genetic variants that had been identified in males as instruments, but conduced, in agreement with the very different production of oestradiol in males and females, that these genetic variants were probably invalid as instruments in females[24]. To our knowledge, no previous

study has been conducted to identify a causal effect of oestradiol on ovarian, breast, or endometrial cancer using multiple instruments for oestradiol levels.

Here, we undertake an MR approach, including four genetic variants that were recently shown to be associated with high oestradiol blood levels in females [26], aiming to establish a causal effect of oestradiol on breast, endometrial and ovarian cancer. We perform a one-sample MR study using data from women in the UK Biobank cohort, as well as a two-sample MR approach using three different independent cancer cohorts, with no overlapping samples with UK Biobank.

Methods

Study samples

UK Biobank

The UK Biobank is a cross-sectional cohort, with both a pro and retrospective study design that includes 502,682 participants, of which 273,404 are women, recruited from all across the United Kingdom (UK). Participants were between 37 and 73 years old at the time of recruitment between 2006 and 2010. Health variables, including cancer diagnosis, have been collected through questionnaires, interviews and hospital records. Participants also answered questions about diet, and lifestyle factors. A total of 820,967 genotyped single nucleotide polymorphisms (SNPs) and up to 90 million imputed variants are available for most participants. For all genetic analyses in UKB the third release of the imputed genetic data was used. The UK Biobank study was approved by the National Research Ethics Committee (REC reference 11/NW/0382). Informed consent to the study was given by all participants. An application for using data from UK Biobank has been approved (application nr: 41143). The UK Biobank analysis performed in this study has also been approved by the Swedish Ethical Review Authority (dnr: 2020-04415).

Oestradiol levels and instrumental variables for oestradiol

Oestradiol levels were measured from blood samples taken at the first assessment in association with the recruitment. Oestradiol was measured by a two-step competitive analysis on a Beckman Coulter Unicel Dxl 800. The instrument variables to be used in the MR analyses were selected from a previous GWAS for oestradiol in UK Biobank [26]. Women with oestradiol levels below 175 pmol/L (detection level) were defined as controls, and

participants with oestrogen above, or equal to175 pmol/L, were defined as cases. Briefly, the GWAS included Caucasian UK Biobank participants clustering with regards to their genetic principal components. First- and second-degree relatives (genetic relationship > 0.044) and samples with sex discordance, high heterozygosity/missingness, and with more than 5% missing genotypes had been excluded. Also, females with unknown menopausal status were removed. After the above filtering, 163,985 females remained in the GWAS of which 37,461 had oestradiol levels above the detection limit (27,463 pre-menopausal and 9,998 post-menopausal). The GWAS was performed using an additive genetic logistic regression model implemented in PLINK 2.0 (2.00-alpha-2-20190429). In total, 4 instrumental variables were identified ($P<1x10^{-7}$) (Table 1). The variance explained by each genetic variant was estimated by calculating the difference between Nagelkerke's pseudo-R^2 for the full model, including both covariates and the SNP, and the reduced model, only including the covariates. The *F*-statistic for each SNP was estimated from the full model by computing the squared ratio of the SNP's beta estimate and its standard error.

Cancer incidences and covariates in UK Biobank

To assess information regarding cancer incidences, data were taken from different categories: main and secondary diagnoses made during a hospital stay, medical conditions assessed both from verbal interviews and touchscreen questionnaires, as well as cause of death and cancer registers. Data taken from the hospital stay, the cancer registry and the death registry were categorised according to the International Classification of Diseases, revision nine (ICD-9) and ten (ICD-10). The ICD10 code C50 (malignant neoplasm of breast) and the ICD9 code 174 (malignant neoplasm of female breast) were used to define breast cancer cases, the ICD10 code C56 (malignant neoplasm of ovary) and the ICD9 code 183 (malignant neoplasm of ovary and other uterine adnexa) were used to define ovarian cancer cases, and the ICD10 code C541 (endometrium) was used to identify endometrial cancer cases. For a detailed description of data-fields and coding used for each cancer and covariate, see Supplementary Table S1-S2. The largest number of cancer cases was identified from the cancer registry data, but since the registry data is sparse before 1995, several additional cancer cases were identified from self-reported data (Supplementary Table S1). General characteristics (Table 1) and information on covariates and potential confounders (Supplementary Table S2) were assessed from data collected during the initial visit to the assessment centre.

To estimate the effect of the oestradiol instrumental variables on the risk for breast, ovarian and endometrial cancer in UK Biobank, we performed a GWAS using a logistic additive genetic model implemented in PLINK 2.0 (2.00-alpha-2-20190429). The following covariates (same as in the oestradiol GWAS) were included: age, BMI, the first ten ancestry derived principal components (to adjust for population structures and ethnic origins), MHT (ever, never and current), oral contraceptive use (ever, never or current), number of live births, menopausal status (pre or post) and hysterectomy (yes or no). In addition, to adjust for the different genotyping chips used in UK Biobank, a binary indicator variable for UK Biobank Axiom versus UK BiLEVE genotyping array was included.

Causal estimation of the effect of oestradiol on breast, endometrial and ovarian cancer

To assess a possible causal effect of high oestradiol levels on the risk of being diagnosed with breast, endometrial or ovarian cancer, we first performed a one-sample MR in UK Biobank. Secondly, we performed a two-sample MR, where the estimates for the oestradiol measurement were extracted from the UK Biobank oestradiol GWAS and the cancer estimates from publicly available GWAS data.

The main MR analyses were performed with the R package gsmr [27] (version 1.0.8), in order to enable causal inference based on multiple SNPs. The HEIDI-outlier flag was disabled as there were too few SNPs (<5) for the procedure to work properly. Note that, all four SNPs are pairwise independent (all located on different chromosomes). We further performed sensitivity analyses using the robust inverse variance weighted (IVW), weighted median, and the MR-Egger methods, included in the "MendelianRandomization" tool implemented in R [28]. The standard IVW method provides a consistent estimate of the causal effect of the exposure on the outcome when all genetic variants are valid instruments [29]. Here, we applied the robust IVW method that is less susceptible to heteroscedasticity and down-weights the effect of possible outlier instruments using robust regression. The two other methods, weighted median and MR-Egger, have been proposed to provide consistent causal estimates for genetic variants under weaker assumptions. The weighted median method only assumes that a majority of the instruments are valid, has a lower bias than the IVW method and owns a reasonable power to detect a causal effect [30]. MR-Egger's intercept test was used to assess the presence of directional pleiotropy [31].

Publicly available GWAS data for two-sample MR analyses

A two-sample MR approach applied for each cancer was done to further support our results from the one-sample MR. In a one-sample MR, any weak-instruments bias is directed towards the confounded association, which could result in inflated MR estimates. In a two-sample MR, this bias is instead directed towards null. Furthermore, the type I error is not inflated in a two-sample setting. In our two-sample MR, the effects of the instruments on cancers were taken from the same oestradiol GWAS as the one-sample MR. However, for breast, endometrial and ovarian cancer respectively (Table 1), the effect sizes and standard errors for the instruments were extracted from publicly available GWAS data, not including UK Biobank participants. For breast cancer, we used summary statistics from the Breast Cancer Association Consortium (BCAC) [32], including 122,977 breast cancer cases and 105,974 healthy controls of European ancestry (downloaded from http://bcac.ccge.medschl.cam.ac.uk/ on 22nd of February 2021). For endometrial cancer, we used data from the Endometrial Cancer Association Consortium (ECAC), which include twelve research cohorts based in Australia, Europe and the USA. From the ECAC cohort, GWAS summary statistics from O'Mara et al (2018) [33], excluding participants from UK Biobank to avoid sample overlap, were used. This restricted ECAC dataset consisted of 12,270 cancer cases and 46,126 controls of European descent. Data from ECAC was available after request from the authors [33]. For ovarian cancer, we used GWAS summary statistics from the Ovarian Cancer Association Consortium (OCAC)[34]. For OCAC, genetic association analysis had been performed for 25,509 epithelial ovarian cancer cases and 40,941 healthy controls. Summary statistics were downloaded from http://ocac.ccge.medschl.cam.ac.uk/data-projects/results-lookup-by-region/, on 26th February 2021.

Data Availability

The data used for this study is available for bona fide researchers from the UK Biobank Resource (http://www.ukbiobank.ac.uk/about-biobank-uk/), and can be accessed by an application to the UK Biobank. The oestradiol data was taken from the supplementary material previously published at https://doi.org/10.5281/zenodo.4926701 [26]. The BCAC data can be downloaded from http://bcac.ccge.medschl.cam.ac.uk/ and the OCAC can be downloaded from http://ocac.ccge.medschl.cam.ac.uk/data-projects/results-lookup-by-region/. The ECAC data was approved after request to the authors. Summary statistics for the published MR study is found in Table 1.

Results

The GWAS for breast, ovarian and endometrial cancer in UK Biobank included 209,877 non-related Caucasian women with complete information for each covariate. In this subset, there were 13,179 breast cancer cases, 1,891 endometrial cancer cases, and 1,477 ovarian cancer cases that were included for each one-sample analysis. We have shown previously that the cancer rate in UK Biobank is similar to the general population in UK for all three cancers studied [18]. The mean age of the UK Biobank subset was 56 years, and most women (74%) had entered menopause. Characteristics for UK Biobank are presented in Table 2. Out of the four instrumental variables selected for oestradiol, two were nominally associated (P<0.05) with breast cancer and one with endometrial cancer in UK Biobank (Table 1). In the BCAC cohort, the SNP rs10638101 was not genotyped, and the proxy rs897797 was therefore included in the two-sample MR for breast cancer. rs897797 was in perfect LD with rs10638101 (R^2=1.0). In addition, one more instrumental variable was associated (P<0.05) with endometrial cancer and one with ovarian cancer (Table 1) in the two-sample cohorts.

Mendelian randomization (MR)

Using our primary MR-method, generalized summary-data-based Mendelian-randomization (GSMR), a significant effect of high oestradiol levels on breast cancer was identified, using both a one-sample approach in UK Biobank (OR=1.30 [95% CI 1.07-1.57]) and a two-sample approach with GWAS summary statistics from BCAC (OR=1.19 [95% CI 1.03-1.38]) (Figure 1, Table 3). Also, for endometrial cancer, we identified a causal effect of oestradiol both in the one-sample MR in UK Biobank (OR=2.01 [95% CI 1.21-3.31]) and in the two-sample approach with cancer estimates from ECAC (OR=1.45 [95% CI 1.14-1.83]) (Figure 1, Table 3). For ovarian cancer, we did not identify a significant effect of oestradiol in the one-sample approach (OR=1.55 [95% CI 0.91-2.65]). Neither did we identify a significant effect in the two-sample approach using GWAS summary statistics from OCAC (OR=1.20 [95% CI 0.99-1.46]) (Figure 1, Table 3).

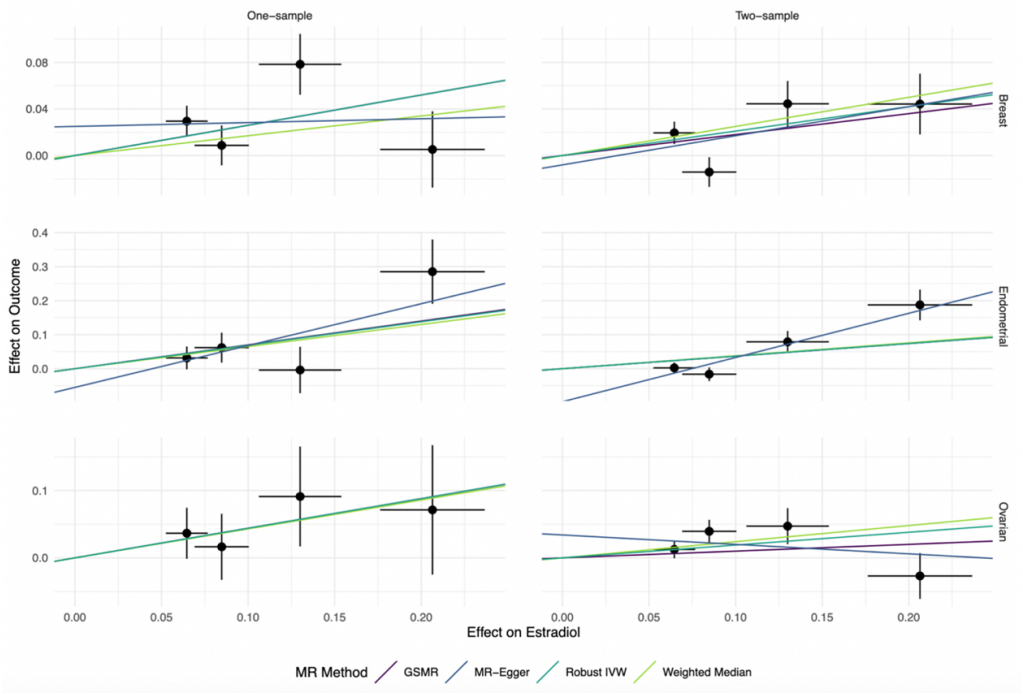

Figure 1. Results from the four different Mendelian randomization methods applied (GSMR, MR-Egger, Robust IVW and Weighted Median) to estimate the causal effect of oestradiol on breast, endometrial and ovarian cancer. The black dots represent the effect size of the SNPs in the GWAS for oestradiol (x-axis) and cancer (y-axis) and the black lines are the standard errors. The lines represent the estimates from the different MR methods.

Sensitivity analyses

As sensitivity analyses, we applied weighted median, robust inverse weighted median and the MR-Egger approach for each cancer, in both the one- and two-sample analyses. Comparing each method, we could see that all three cancers showed the same direction of effect with all methods (Figure 1, Table 3). The MR-Egger intercept for endometrial cancer was significant in the two-sample analysis, which indicates that we might have directional pleiotropy. However, the MR-Egger approach controls for this type of pleiotropy given that the instrument strength independent of direct effect assumption is not violated [31]. Since the MR-Egger beta estimate was still significant with the same direction of effect as in all other methods, we therefore argue that the results from MR-Egger are consistent with the other methods.

Discussion

In this study, we have used a one- and a two-sample MR approach to show that high levels of oestradiol in the body, promotes the development of breast and endometrial cancer.

In a previous MR study, including 6,608 endometrial cancer cases and 37,925 postmenopausal controls, using one single instrument (rs727479) mapped to the *CYP19A1* gene, it was suggested that oestradiol has a causal effect on endometrial cancer risk, but no significant effect was identified for breast cancer [23]. Beside that the previous study was smaller than ours, and might not have had enough power to identify a causal effect for breast cancer, their selected instrument might have been less appropriate to use since it appears to be more strongly associated with oestradiol in males than in females[26], which would bias the causal effect towards null. In addition, any bias due to horizontal pleiotropy may go undetected, by using a single instrument. Another, recent study by Larsson *et al*[24], identified an effect of oestradiol on both ER-positive breast cancer and endometrial cancer, using the same genetic variant, also with the effect on oestradiol being estimated in males. Even though rs727479 was not available in UK Biobank, it is in perfect LD with rs7175531 (R^2=1.0), which we previously showed to be strongly associated with oestradiol in male UK Biobank participants (P=1.32x10^{-60}) [26]. However, in UK Biobank, rs7175531 was only nominally significant, in pre- and post-menopausal women (P=0.020 and P=0.043, respectively). rs7175531 is intronic in *CYP19A1*, which encodes an aromatase that synthesizes endogenous oestrogens from testosterone in adipose tissue [35]. This suggests that this locus might be more important for oestradiol production in males, and not play an as important role in the oestrogen production in females. The *CYP19A1* SNP could thereby be an invalid MR instrument for estimating the causal effect of oestradiol on female-specific phenotypes.

In this MR study, we instead used four oestradiol instruments (SNPs), that have been associated with oestradiol levels in females [26]. These SNPs were annotated to *CYP3A7, ASCL1, TMEM150B* and *MCM8*. The most significant association to high oestradiol levels was identified for the *CYP3A7* SNP (P=7.62x10^{-12}). *CYP3A7* encodes cytochrome P450 CYP3A7, which metabolizes dehydroepiandrosterone (DHEA), which is the main precursor of circulating oestrogens in women [36,37]. The instrument for *CYP3A7* (rs45446698) was nominally significant (P<0.05) for endometrial cancer in both UK Biobank and ECAC, and for breast cancer in BCAC (Table 1). CYP3A7 is mainly expressed in the liver, which is one

of the primary sites of oestrogen metabolism [38]. CYP3A7 is also expressed in the endometrium [39]. The second most significant oestradiol instrument was the *MCM8* SNP, rs16991615 that has previously been shown to be one of the strongest associated loci also for age at menopause [40]. This instrument was nominally significant (P<0.05) for breast cancer both in UK Biobank (P=0.0027) and in BCAC (P=0.024), as well as for endometrial cancer in ECAC (P=0.013) (Table 1). Even though the previous oestradiol GWAS did not adjust for age at menopause, a significant difference in the estimate was not identified when including postmenopausal women only and adjusting for age at menopause [26]. This suggests that this effect is not confounded by age at menopause. An oestradiol GWAS excluding all participants with a cancer diagnosis prior to assessment (when blood samples were drawn), was also performed as a sensitivity analysis. This analysis did not show any significant difference in effect for the instruments, indicating that a prior cancer diagnosis did not influence the results [26]. *ASCL1,* the gene mapping to our third most significant oestradiol instrument, has previously been shown to promote tumour progression in lung adenocarcinoma [41] and overall survival in ovarian cancer patients [42]. The instrument mapped to *ASCL1* was nominally significant for ovarian cancer within the OCAC cohort (P=0.021, Table 1). *TMEM150B* has, except for being associated with age at menopause [43], also been associated with age at menarche [44]. The instrument mapping to *TMEM150B* (rs10638101) was nominally significant for breast cancer in UK Biobank (P=0.026), and its proxy (rs897797) was also nominally significant for breast cancer in BCAC (P=0.043; Table 1).

Even though we used a larger number of instrumental variables in our MR analyses than previous oestradiol MR studies on cancer risk, one of the main limitations of our study is still that the number of instruments is limited. In order for the HEIDI outlier procedure in the GMSR method to work the best, at least ten instrumental variables are needed. This means that we were not able to adjust for pleiotropy by removing instruments that are identified as pleiotropic. However, today, no other studies, with larger sample sizes than UK Biobank exist, and the identification of additional instruments is therefore not possible at this time. For the two-sample MR analyses, the MR-Egger intercept was significantly below zero for endometrial cancer, which could indicate that we have problems with pleiotropy and the results from the current study would benefit from replication, once additional instrumental variables for oestradiol are available. MR-Egger is not appropriate to run for a one-sample approach, and we, therefore, should interpret the MR-Egger results for the one-sample analysis with care [45]. Another limitation of this study is the hormonal fluctuations during the

menstrual cycle and that oestradiol is commonly measured at different time points during the menstrual cycle in different women, but also that the oestradiol measured in blood is the result of the oestrogen produced in the ovaries (in pre-menopausal women), together with the oestradiol produced in other tissues. More detailed measurements of oestradiol during different time points of the menstrual cycle would be beneficial to address the causal effects of oestradiol pre-menopause. Since oestradiol mainly is produced by the ovaries during the reproductive years, and mainly by subcutaneous adipose tissues after menopause [5], another limitation is therefore that the oestradiol GWAS and MR includes both pre- and post-menopausal women.

With regards to very few oestradiol instruments associated with ovarian cancer, it is not surprising that we did not find a causal relationship between high oestradiol levels and risk of ovarian cancer. However, the MR estimates from our one- and two-sample MR are both in the same direction, and at least the two-sample MR would have met the significance thresholds of a one-sided test, testing only if high oestradiol levels increase the risk of ovarian cancer rather than it influencing the risk in any direction (increase or decrease). Therefore, we cannot rule out the possibility of a true causal effect that might be identified in a larger set of ovarian cancer cases. However, the associated effect of oestrogen on the risk of ovarian cancer is not as strong [46] as that previously seen for breast [47], and especially endometrial cancer [4], which agrees with our results. On the other hand, ovarian cancer is a heterogeneous disease with various subtypes which are different concerning their etiology, morphology, biology, pathogenicity and prognosis [48], therefore, future studies should benefit from include larger cohorts and focus on more distinct subtypes.

Although oestradiol is involved in regulating different features of human physiology in both genders, the underlying mechanisms through which oestrogen exerts its carcinogenic effects are not yet fully established and different hypotheses have been postulated so far. One hypothesis is that oestradiol has toxic effects and promotes mutations in the DNA [49] or that the oestrogen metabolites can cause genomic instability and mutations that can result in malignancy [50,51]. However, it has also been proposed that oestradiol can restrain the DNA damage response, which is a cellular mechanism to repair DNA damage, and thereby, may play a further role in the development and progression of the neoplasms [49]. Another hypothesis is that the cancerogenic effects of oestradiol mainly take place through binding to one of its two receptors (i.e., oestrogen receptor (ER) α and β). While the oestradiol-ERβ complex inhibits

cell proliferation, the binding of the hormone to ERα in the oestradiol-ERα complex induces cell proliferation, which has been suggested as a mechanism of the neoplastic effect of oestradiol [2,4]. In agreement with this, it has been shown that oestrogen signalling via ERα induces hyperproliferation and carcinogenic transformation in breast, ovarian and endometrial cells [5,52].

By identifying and replicating a possible causal link between oestradiol and endometrial, as well as for breast cancer, using a more valid approach compared to previous research, our results further support carcinogenic effects of oestrogen in these tissues. A deeper understanding of such possibly causal relation is also of importance for novel interventions to prevent cancer in women.

Author contributions

WEE and ÅJ designed the study; WEE performed the MR analysis and cancer GWASes, DS generated the figures and performed the oestradiol GWAS; WEE and TK performed the statistical analysis; WEE, TK, JH, DS, ÅJ, FH wrote the manuscript; DS, JH, TK, WEE, FH and ÅJ interpreted the data, contributed to and reviewed the manuscript. All authors declare no conflicts of interest.

Acknowledgements

We acknowledge participants and staff at UK Biobank. The computations were performed on resources provided by SNIC through Uppsala Multidisciplinary Centre for Advanced Computational Science (UPPMAX) under project sens2017538. The breast cancer genome-wide association analyses were supported by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research, the 'Ministère de l'Économie, de la Science et de l'Innovation du Québec' through Genome Québec and grant PSR-SIIRI-701, The National Institutes of Health (U19 CA148065, X01HG007492), Cancer Research UK (C1287/A10118, C1287/A16563, C1287/A10710) and The European Union (HEALTH-F2-2009-223175 and H2020 633784 and 634935). All studies and funders are listed in Michailidou et al (Nature, 2017). The endometrial cancer genome-wide association analyses were supported by the National Health and Medical Research Council of Australia (APP552402, APP1031333, APP1109286, APP1111246 and APP1061779, APP1173170),

the U.S. National Institutes of Health (R01-CA134958), European Research Council (EU FP7 Grant), Wellcome Trust Centre for Human Genetics (090532/Z/09Z) and Cancer Research UK. OncoArray genotyping of ECAC cases was performed with the generous assistance of the Ovarian Cancer Association Consortium (OCAC), which was funded through grants from the U.S. National Institutes of Health (CA1X01HG007491-01 (C.I. Amos), U19-CA148112 (T.A. Sellers), R01-CA149429 (C.M. Phelan) and R01-CA058598 (M.T. Goodman); Canadian Institutes of Health Research (MOP-86727 (L.E. Kelemen)) and the Ovarian Cancer Research Fund (A. Berchuck). We particularly thank the efforts of Cathy Phelan. OncoArray genotyping of the BCAC controls was funded by Genome Canada Grant GPH-129344, NIH Grant U19 CA148065, and Cancer UK Grant C1287/A16563. All studies and funders included in ECAC are listed in O'Mara et al (2018). We also like to acknowledge Tracy O'Mara at QIMR Berghofer for sharing the ECAC data, after removing the UK Biobank.

References

1. Wise, P. M., Suzuki, S. & Brown, C. M. Estradiol: A hormone with diverse and contradictory neuroprotective actions. *Dialogues in Clinical Neuroscience* **11**, 297–303 (2009).
2. Ascenzi, P., Bocedi, A. & Marino, M. Structure-function relationship of estrogen receptor α and β: Impact on human health. *Struct. Relatsh. estrogen Recept. α β Impact Hum. Heal.* 299–402 (2006).
3. Hilton, H. N., Clarke, C. L. & Graham, J. D. Estrogen and progesterone signalling in the normal breast and its implications for cancer development. *Molecular and Cellular Endocrinology* **466**, 2–14 (2018).
4. Rodriguez, A. C., Blanchard, Z., Maurer, K. A. & Gertz, J. Estrogen Signaling in Endometrial Cancer: a Key Oncogenic Pathway with Several Open Questions. *Horm. cancer* **10**, 51–63 (2019).
5. Zhao, H., Zhou, L., Shangguan, A. & Bulun, S. Aromatase expression and regulation in breast and endometrial cancer. *J. Mol. Endocrinol.* **57**, 19–33 (2016).
6. Van Weelden, W., Massuger, L. & Pijnenborg, J. Anti-estrogen treatment in endometrial cancer: A systematic review. *Front. Oncol.* **9**, 359 (2018).
7. Salehi, F., Dunfield, L., Phillips, K. P., Krewski, D. & Vanderhyden, B. C. Risk factors for ovarian cancer: An overview with emphasis on hormonal factors. *Journal of Toxicology and Environmental Health - Part B: Critical Reviews* **11**, 301–321 (2008).
8. Groothuis, P., Dassen, H., Romano, A. & Punyadeera, C. Estrogen and the endometrium: Lessons learned from gene expression profiling in rodents and human. *Hum. Reprod. Updat. Oxford Acad.* **13**, 405–417 (2007).
9. Reed, B. & Carr, B. The Normal Menstrual Cycle and the Control of Ovulation. *Endotext. MDText.com, Inc* (2000).
10. Anderson, E., Clarke, R. B. & Howell, A. *Estrogen Responsiveness and Control of Normal Human Breast Proliferation*. *Journal of Mammary Gland Biology and Neoplasia* **3**, 23–35 (Kluwer Academic/Plenum Publishers, 1998).
11. Garvin, S., Nilsson, U. W., Huss, F. R. M., Kratz, G. & Dabrosin, C. Estradiol increases VEGF in human breast studied by whole-tissue culture. *Cell Tissue Res.* **325**, 245–251 (2006).
12. Group, E. H. and B. C. C. *et al.* Sex hormones and risk of breast cancer in premenopausal women: A collaborative reanalysis of individual participant data from

seven prospective studies. *Lancet Oncol.* **14**, 1009–1019 (2013).

13. Kaaks, R. *et al.* Serum sex steroids in premenopausal women and breast cancer risk within the European Prospective Investigation into Cancer and Nutrition (EPIC). *J. Natl. Cancer Inst.* **97**, 755–765 (2005).
14. Zhang, X., Tworoger, S. S., Eliassen, A. H. & Hankinson, S. E. Postmenopausal plasma sex hormone levels and breast cancer risk over 20 years of follow-up. *Breast Cancer Res. Treat.* **137**, 883–892 (2013).
15. Kaaks, R. *et al.* Postmenopausal serum androgens, oestrogens and breast cancer risk: The European prospective investigation into cancer and nutrition. *Endocr. Relat. Cancer* **12**, 1071–1082 (2005).
16. Brinton, L. A. & Felix, A. S. Menopausal hormone therapy and risk of endometrial cancer. *J. Steroid Biochem. Mol. Biol.* **142**, 83–89 (2014).
17. Mungenast, F. & Thalhammer, T. Estrogen biosynthesis and action in ovarian cancer. *Front. Endocrinol. (Lausanne).* **5**, (2014).
18. Karlsson, T., Johansson, T., Hoglund, J., Ek, W. E. & Johansson, Å. Time-dependent effects of oral contraceptive use on breast, ovarian, and endometrial cancers. *Cancer Res.* **81**, 1153–1162 (2021).
19. Iversen, L., Sivasubramaniam, S., Lee, A. J., Fielding, S. & Hannaford, P. C. Lifetime cancer risk and combined oral contraceptives: the Royal College of General Practitioners' Oral Contraception Study. *Am. J. Obstet. Gynecol.* **216**, 580.e1-580.e9 (2017).
20. Berchuck, A. & Scildkraut, J. Oral contraceptive pills. Prevention of ovarian cancer and other benefits. *N C Med J.* **58**, 405–407 (1997).
21. Stewart, C., Raylea, C. & Lockwood, S. Ovarian Cancer: An Integrated Review. Seminars in Oncology Nursing. W.B. *Saunders* **35**, 151–156 (2019).
22. Didelez, V. & Sheehan, N. Mendelian randomization as an instrumental variable approach to causal inference. *Stat. Methods Med. Res.* **16**, 309–330 (2007).
23. Thompson, D. J. *et al.* CYP19A1 fine-mapping and Mendelian randomization: estradiol is causal for endometrial cancer. *Endocr. Relat. Cancer* **23**, 77–91 (2016).
24. Larsson, S. C. *et al.* Serum Estradiol and 20 Site-Specific Cancers in Women: Mendelian Randomization Study. *J. Clin. Endocrinol. Metab.* (2021).
25. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).
26. Schmitz, D. *et al.* Genome-Wide Association Study of Estradiol Levels, and the Causal

Effect of Estradiol on Bone Mineral Density. *JCEM* (2021). doi:10.1101/2021.07.01.21259826

27. Zhu, Z. *et al.* Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat. Commun.* **9**, (2018).
28. Olena, Y. O. & Burgess, S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int. J. Epidemiol.* **46**, 1734–1739 (2017).
29. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian Randomization Analysis With Multiple Genetic Variants Using Summarized Data. *Genet. Epidemiol.* **37**, 658–665 (2013).
30. Bowden, J., Smith, D. G. & Burgess, S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Int. J. Epidemiol.* **44**, 512–525 (2015).
31. Bowden, J., Smith, G. D. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–525 (2015).
32. Michailidou, K. *et al.* Association analysis identifies 65 new breast cancer risk loci. *Nature* **551**, 92–94 (2017).
33. O'Mara, T. A. *et al.* Identification of nine new susceptibility loci for endometrial cancer. *Nat. Commun.* **9**, (2018).
34. Phelan, C. M. *et al.* Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. *Nat. Genet.* **49**, 680–691 (2017).
35. Flote, V. G. *et al.* Gene variations in oestrogen pathways, CYP19A1, daily 17β-estradiol and mammographic density phenotypes in premenopausal women. *Breast Cancer Res.* **16**, (2014).
36. Lee, A. J., Conney, A. H. & Zhu, B. T. Human Cytochrome P450 3A7 Has a Distinct High Catalytic Activity for the 16α-Hydroxylation of Estrone but not 17β-Estradiol. *Cancer Res.* **63**, 6532 LP – 6536 (2003).
37. Ohmori, S. *et al.* Differential catalytic properties in metabolism of endogenous and exogenous substrates among CYP3A enzymes expressed in COS-7 cells. *Biochim. Biophys. Acta - Gen. Subj.* **1380**, 297–304 (1998).
38. Tsuchiya, Y., Nakajima, M. & Yokoi, T. Cytochrome P450-mediated metabolism of estrogens and its regulation in human. *Cancer Lett.* **227**, 115–124 (2004).
39. Schuetz, J. D., Kauma, S. & Guzelian, P. S. Identification of the fetal liver cytochrome

CYP3A7 in human endometrium and placenta. *J. Clin. Invest.* **92**, 1018–1024 (1993).

40. Chen, C. T. L. *et al.* Meta-analysis of loci associated with age at natural menopause in African-American women. *Hum. Mol. Genet.* **23**, 3327–3342 (2014).
41. Miyashita, N. *et al.* ASCL1 promotes tumor progression through cell-autonomous signaling and immune modulation in a subset of lung adenocarcinoma. *Cancer Lett.* **489**, 121–132 (2020).
42. Moore, K. N. *et al.* Genome-wide association study evaluating single-nucleotide polymorphisms and outcomes in patients with advanced stage serous ovarian or primary peritoneal cancer: An NRG Oncology/Gynecologic Oncology Group study. *Gynecol. Oncol.* **147**, 396–401 (2017).
43. Stolk, L. *et al.* Meta-analyses identify 13 loci associated with age at menopause and highlight DNA repair and immune pathways. *Nat. Genet.* **44**, 260–268 (2012).
44. Pickrell, J. K. *et al.* Detection and interpretation of shared genetic influences on 42 human traits. *Nat. Genet.* **48**, 709–717 (2016).
45. Bowden, J. Misconceptions on the use of MR-Egger regression and the evaluation of the InSIDE assumption. *Int. J. Epidemiol.* **46**, 2097–2099 (2017).
46. Trabert, B. *et al.* Circulating estrogens and postmenopausal ovarian cancer risk in the women's health initiative observational study. *Cancer Epidemiol. Biomarkers Prev.* **4**, 648–656 (2016).
47. Key, T., Appleby, P., Hines, L. & Al., E. Circulating sex hormones and breast cancer risk factors in postmenopausal women: reanalysis of 13 studies. *Br J Cancer* **105**, 709–722 (2011).
48. Kossai, M., Leary, A., Scoazec, J.-Y. & Genestine, C. Ovarian Cancer: A Heterogeneous Disease. *Karger Publ.* **85**, 41–49 (2018).
49. Pescatori, S. *et al.* A Tale of Ice and Fire: The Dual Role for 17β-Estradiol in Balancing DNA Damage and Genome Integrity. *Cancers (Basel).* **13**, (2021).
50. Cavalieri, E., Rogan, E. & Zahid, M. Critical depurinating DNA adducts: Estrogen adducts in the etiology and prevention of cancer and dopamine adducts in the etiology and prevention of Parkinson's disease. *Int. J. Cancer* **141**, (2017).
51. Wen, C., Wu, L., Fu, L., Wang, B. & Zhou, H. Unifying mechanism in the initiation of breast cancer by metabolism of estrogen (Review). *Molecular Medicine Reports* **16**, 1001–1006 (2017).
52. Li, H. H. *et al.* Estradiol 17β and its metabolites stimulate cell proliferation and antagonize ascorbic acid-suppressed cell proliferation in human ovarian cancer cells.

Reprod. Sci. **21**, 102–111 (2014).

Tables

Table 1. Summary GWAS results for each instrument variable included in the MR analysis. BCAC; Breast Cancer Association Consortium, OCAC; Ovarian Cancer Association Consortium, ECAC; Endometrial Cancer Association Consortium. Significant values (P<0.05) are highlighted in bold.

						Breast Cancer				**Ovarian Cancer**				**Endometrial Cancer**			
SNP	**Effect allele/ freq***	**Chromosome position: base pair**	**Delta R2*** / F-statistics**	**Oestradiol Levels OR (95% CI)**	**P-value**	**UK Biobank**		**BCAC**		**UK Biobank**		**OCAC**		**UK Biobank**		**ECAC**	
						OR (95% CI)	**P-value**	**OR (95% CI)**	**P-value**	**OR (95% CI)**	**P-value**	**OR (95% CI)**	**P-value**	**OR (95% CI)**	**P-value**	**OR (95% CI)**	**P-value**
rs45446698	T/0.96	Chr7: 99332948	0.00026 / 45.16	1.23 (1.16-1.30)	7.62×10^{-12}	1.01 (0.94-1.07)	0.87	1.05 (0.99-1.10)	0.09	1.07 (0.89-1.30)	0.51	0.97 (0.91-1.04)	0.43	**1.33 (1.11-3.02)**	**0.0037**	**1.21 (1.10-1.32)**	**3.09×10^{-5}**
rs4764934	C/0.82	Chr12: 103469376	0.00015 / 26.42	1.09 (1.06-1.12)	6.07×10^{-8}	1.01 (0.96-1.04)	0.61	0.99 (0.96-1.01)	0.26	1.02 (0.92-1.12)	0.74	**1.04 (1.01-1.08)**	**0.021**	1.06 (0.98-2.65)	0.16	0.98 (0.95-1.02)	0.42
rs10638101	A/0.51	Chr19: 55827728	0.00016 / 28.20	1.07 (1.04-1.09)	6.28×10^{-8}	**1.03 (1.00-1.06)**	**0.026**	Missing		1.04 (0.96-1.12)	0.33	1.01 (0.99-1.04)	0.34	1.03 (0.97-1.10)	0.34	1.00 (0.97-1.03)	0.88
rs897797**	T/0.50	Chr19: 55831723	0.00016 / 28.52	1.07 (1.04-1.09)	7.89×10^{-8}			**1.02 (1.00-1.04)**	**0.043**								
rs16991615	A/0.07	Chr20: 5948227	0.00016 / 28.94	1.14 (1.09-1.19)	4.67×10^{-8}	**1.08 (1.08-1.09)**	**0.0027**	**1.05 (1.01-1.09**	**0.024**	1.10 (0.95-1.27)	0.22	1.05 (0.96-1.15)	0.078	1.00 (0.87-1.14)	0.95	**1.08 (1.02-1.15)**	**0.013**

* Allele frequency for effect allele

** Proxy SNP, included since rs10638101 was not genotyped in BCAC

*** Delta R^2 denotes the difference in Nagelkerke's pseudo-R^2 between the full model, including both covariates and SNP, and the reduced model, only including covariates.

Table 2. Characteristics in the UK Biobank participants

	Breast Cancer		**Ovarian Cancer**		**Endometrial Cancer**	
	Cases	**Controls**	**Cases**	**Controls**	**Cases**	**Controls**
Participants, N (%)	13179 (6.77)	181628	1477 (0.71)	193330	1891 (0.97)	192916
Entered menopause (%)	11200 (84.98)	131990 (72.67)	1313 (88.90)	141877 (73.39)	1719 (90.90)	141471 (73.33)
BMI*, mean (Q1 – Q3)	27.19 (23.74-29.80)	27.02 (23.42-29.80)	27.60 (23.94-30.25)	27.03 (23.44-29.63)	29.75 (24.92-33.44)	27.00 (23.43-29.60)
Age, mean (Q1 – Q3)	59.19 (55-61)	56.50 (50-63)	59.28 (54-65)	56.66 (51-63)	60.75 (58-65)	56.64 (50-71)
Smoking N (ever/never)	7472/5657	107989/73032	637/835	78052/114626	1145/737	114316/21150
Had hysterectomy, N (%)	2882 (21.87)	34214 (18.83)	816 (55.25)	36280 (18.77)	1010 (53.41)	36086 (18.71)
Had menopausal hormone therapy, N (%)	5462 (41.44)	17415 (9.59)	783 (53.01)	76094 (39.36)	868 (45.90)	76009 (39.40)
Oral contraceptives, N ever users (%)	10412 (79.00)	150074 (82.63)	1083 (73.32)	159403 (82.45)	1275 (67.42)	159211 (82.53)
Number of live births, median (Q1 – Q3)	2 (1-2)	2 (1-2)	2 (1-2)	2 (1-2)	2 (1-2)	2 (1-2)
Age at menopause, median (Q1 – Q3)	46.12 (46-52)	46.49 (46-53)	45.59 (45-52)	46.47 (46-53)	46.64 (48-52)	46.46 (46-53)

*Body mass index

Table 3. Mendelian Randomisation results for each cancer and method. BCAC; Breast Cancer Association Consortium, ECAC; Endometrial Cancer Association Consortium, OCAC; Ovarian Cancer Association Consortium.

	Breast Cancer				**Ovarian Cancer**				**Endometrial Cancer**			
	UK Biobank		**BCAC**		**UK Biobank**		**OCAC**		**UK Biobank**		**ECAC**	
Method	OR (95% CI)	P-value	OR (95% CI)	P-value	OR (95% CI)	P-value	OR (95% CI)	P-value	OR (95% CI)	P-value	OR (95% CI)	P-value
GSMR	1.30 (1.07-1.57)	0.0074	1.19 (1.03-1.38)	0.018	1.55 (0.91-2.65)	0.11	1.20 (0.99-1.46)	0.066	2.01 (1.21-3.31)	0.0065	1.45 (1.14-1.83)	0.0022
Weighted Median	1.18 (0.92-1.51)	0.18	1.28 (1.07-1.54)	0.006	1.54 (0.83-2.86)	0.17	1.26 (0.97-1.64)	0.080	1.91 (0.99-3.68)	0.051	1.46 (1.06-2.02)	0.021
Robust IVW	1.30 (1.00-1.68)	0.051	1.24 (0.92-1.67)	0.15	1.55 (1.11-2.16)	0.009	1.21 (0.87-1.69)	0.25	2.00 (1.13-3.53)	0.017	1.44 (0.89-2.33)	0.133
MR-Egger	1.03 (0.49-2.19)	0.93	1.28 (0.66-2.58)	0.46	1.49 (0.40-5.58)	0.55	0.87 (0.45-1.70)	0.69	3.41 (0.65-17.96)	0.15	3.68 (1.73-7.85)	0.001
MR-Egger intercept	1.03 (0.95-1.10)	0.52	0.99 (0.93-1.06)	0.82	1.00 (0.88-1.14)	0.95	1.03 (0.97-1.10)	0.31	0.95 (0.81-1.11)	0.50	0.91 (0.84-0.98)	0.009