Evaluating causal associations between previously reported risk factors and epithelial ovarian cancer: a Mendelian randomization analysis

Short title: Causal appraisal of reported risk factors in epithelial ovarian cancer

James Yarmolinsky, MSc[1,2], Caroline L Relton, PhD*[1,2, 6], Artitaya Lophatananon, PhD[3], Kenneth Muir, PhD[3], Usha Menon, MD[4], Aleksandra Gentry-Maharaj, PhD[4], Axel Walther, PhD FRCP[5], Jie Zheng, PhD[1,2], Peter Fasching, MD[7], Wei Zheng, PhD[8], Woo Yin Ling, PhD[9], Jenny Chang-Claude, PhD[10,11], Sue K Park, PhD[12,13,14], Byoung-Gie Kim, PhD[15], Ji-Yeob Choi, PhD[12,13,14], Boyoung Park, PhD[16], George Davey Smith, DSc[1,2,6], Richard M Martin, BM BS, PhD*[1,2,6], Sarah J Lewis, PhD*[1,2]

[1] MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK

[2] Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

[3] Division of Population Health, Health Services Research and Primary Care, School of Health Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, United Kingdom

[4] MRC Clinical Trials Unit, Institute for Clinical Trials and Methodology, University College London, London, UK

[5] Bristol Cancer Institute, University Hospitals Bristol NHS Foundation Trust, Bristol, UK

[6] National Institute for Health Research Bristol Biomedical Research Centre, University of Bristol and University Hospitals Bristol NHS Foundation Trust, Bristol, UK

[7] Department of Gynecology and Obstetrics, University Hospital Erlangen, Comprehensive Cancer Center Erlangen-EMN, Friedrich-Alexander University Erlangen-EMN, Erlangen, Germany

[8]Division of Epidemiology, Vanderbilt University Medical Center, Vanderbilt University, Nashville, USA

[9]Faculty of Medicine, University of Malaya, Malaysia

[10]German Cancer Research Center (DKFZ), Division of Cancer Epidemiology, Heidelberg, Germany;

[11]University Medical Center Hamburg-Eppendorf, University Cancer Center Hamburg (UCCH), Cancer Epidemiology Group, Hamburg, Germany

[12]Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, Korea

[13]Cancer Research Institute, Seoul National University, Seoul, Korea

[14]Department of Biomedical Science, Seoul National University Graduate School, Seoul, Korea

[15]Department of Obstetrics and Gynecology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, South Korea

[16]Department of Medicine, College of Medicine, Hanyang University, Seoul, South Korea

*Contributed equally

Corresponding author:

Sarah J Lewis

MRC Integrative Epidemiology Unit,

Population Health Sciences, University of Bristol,

Oakfield House, Oakfield Grove,

Bristol BS8 2BN, UK

Phone: +44 (0) 117 33 13316

E-mail: S.J.Lewis@bristol.ac.uk

Word count: 3522

Abstract

Background

Various modifiable risk factors have been associated with epithelial ovarian cancer risk in observational epidemiological studies. However, the causal nature of the risk factors reported, and thus their suitability as effective intervention targets, is unclear given the susceptibility of conventional observational designs to residual confounding and reverse causation. Mendelian randomization uses genetic variants as proxies for modifiable risk factors to strengthen causal inference in observational studies. We used Mendelian randomization to evaluate the causal role of 13 previously reported risk factors (reproductive, anthropometric, clinical, lifestyle, and molecular factors) in overall and histotype-specific epithelial ovarian cancer in up to 25,509 case subjects and 40,941 controls in the Ovarian Cancer Association Consortium.

Methods and Findings

Genetic instruments to proxy 13 risk factors were constructed by identifying single nucleotide polymorphisms (SNPs) robustly ($P<5\times10^{-8}$) and independently associated with each respective risk factor in previously reported genome-wide association studies. SNPs were combined into multi-allelic inverse-variance weighted fixed or random-effects models to generate causal estimates. Three complementary sensitivity analyses were performed to examine violations of Mendelian randomization assumptions: MR-Egger regression and weighted median and mode estimators. A Bonferroni-corrected *P*-value threshold was used to establish “strong evidence” ($P<0.0038$) and “suggestive evidence” ($0.0038<P<0.05$) for associations.

In Mendelian randomization analyses, there was strong or suggestive evidence that 9 of 13 risk factors had a causal effect on overall or histotype-specific epithelial ovarian cancer.

There was strong evidence that genetic liability to endometriosis increased risk of epithelial ovarian cancer (OR per log odds higher liability:1.27, 95% CI: 1.16-1.40; P=6.94x10^{-7}) and suggestive evidence that lifetime smoking exposure increased risk of epithelial ovarian cancer (OR per unit increase in smoking score:1.36, 95% CI: 1.04-1.78; P=0.02). In histotype-stratified analyses, the strongest associations found were between: height and clear cell carcinoma (OR per SD increase:1.36, 95% CI: 1.15-1.61; P=0.0003); age at natural menopause and endometrioid carcinoma (OR per year later onset:1.09, 95% CI: 1.02-1.16; P=0.007); and genetic liability to polycystic ovary syndrome and endometrioid carcinoma (OR per log odds higher liability:0.74, 95% CI :0.62-0.90; P=0.002). There was little evidence for an effect of genetic liability to type 2 diabetes, parity, or circulating levels of 25-hydroxyvitamin D and sex hormone-binding globulin on ovarian cancer or its subtypes. The primary limitations of this analysis include: modest statistical power for analyses of risk factors in relation to some less common ovarian cancer histotypes (low grade serous, mucinous, and clear cell carcinomas), the inability to directly examine the causal effects of some ovarian cancer risk factors that did not have robust genetic variants available to serve as proxies (e.g., oral contraceptives, hormone replacement therapy), and the assumption of linear relationships between risk factors and ovarian cancer risk.

Conclusions

Our comprehensive examination of possible etiological drivers of ovarian carcinogenesis using germline genetic variants to proxy risk factors supports a causal role for few of these factors in epithelial ovarian cancer and suggests distinct etiologies across histotypes. The identification of novel modifiable risk factors remains an important priority for the control of epithelial ovarian cancer.

Introduction

Ovarian cancer is the second most common gynecological cancer in the USA and Western Europe and accounts for more deaths than all other gynecological cancers combined [1,2]. The prognosis for ovarian cancer is generally poor because women typically present with advanced disease due to the non-specific nature of symptoms and because of the lack of established screening tests [3-5]. Given the limited success of secondary prevention strategies and the sporadic nature of 90% of cases, primary prevention of ovarian cancer may serve as an important vehicle for disease control [6]. However, few modifiable risk factors have consistently been linked to ovarian cancer in observational epidemiological studies and most previous studies have failed to stratify analyses across clinically distinct histotypes [7-10]. Further, the causal nature of the risk factors reported, and thus their suitability as effective intervention targets, is unclear given the susceptibility of conventional observational designs to residual confounding and reverse causation.

Mendelian randomization (MR) is an analytical approach that uses germline genetic variants as instruments ("proxies") for potentially modifiable risk factors, to examine the causal effects of these factors on disease outcomes in observational settings [11,12]. Since germline genetic variants are randomly assorted at meiosis, MR analyses should be less prone to confounding by lifestyle and environmental factors than conventional observational studies. Further, since germline genetic variants are fixed at conception and cannot be influenced by subsequent disease processes, MR analyses are not subject to reverse causation bias. An additional advantage of MR is that it can be implemented using summary genetic association data from two independent samples, representing: a) the genetic variant-risk factor associations; and b) the genetic variant-outcome associations ("two-sample Mendelian randomization"). This provides an efficient and statistically robust method of appraising causal relationships between risk factors and disease outcomes.

Given the current poor understanding of the etiology of epithelial ovarian cancer (EOC), a two-sample Mendelian randomization analysis was performed to evaluate the causal effects of 13 previously reported factors with risk of overall and histotype-specific EOC.

Methods

Ovarian cancer population

Summary genetic association data were obtained on 25,509 women with EOC and 40,941 controls of European descent. These women had been genotyped using the Illumina Custom Infinium array (OncoArray) as part of the Ovarian Cancer Association Consortium (OCAC) genome-wide association study (GWAS) [13,14]. The data included the following invasive ovarian cancer histotypes: high grade serous carcinoma (n=13,037), low grade serous carcinoma (n=1,012), mucinous carcinoma (n=1,417), endometrioid carcinoma (n=2,810), and clear cell carcinoma (n=1,366). Analyses were also performed for low malignant potential tumors (n=3,103) which included 1,954 serous and 1,140 mucinous tumors. Invasive histotypes classified as “other” (n=2,764 cases) were included in analyses for overall epithelial ovarian cancer but were not assessed separately. Ethical approval from relevant research ethics committees was obtained for all studies in OCAC and written, informed consent was obtained from all participants in these studies. Further details about the OCAC study and OncoArray analyses are available in **Supplemental Materials**.

Identification of previously reported risk factors and instrument selection

Previously reported risk factors for EOC were identified from a literature review of narrative and systematic review articles summarizing findings from observational epidemiological studies using PubMed and Web of Science [15-20] and through consultation with the Cancer Research UK website and the World Cancer Research Fund/American Institute for Cancer Research Ovarian Cancer 2014 Report (accessed on 02/10/2017). Genetic instruments were then identified for these risk factors by consulting the preprint server bioRxiv (http://www.biorxiv.org/) and two catalogues of summary GWAS data: the NHGRI-EBI (National Human Genome Research Institute - European Bioinformatics Institute) GWAS catalogue and MR-Base [21,22]. The complete PubMed and Web of Science search strategies and instrument selection criteria are presented in **Supplementary Materials** and **Extended Methods**, respectively.

In total, 13 risk factors with a suitable genetic instrument were included in the analysis: four reproductive factors (age at menarche, age at natural menopause, parity, and genetic liability to twin pregnancy)[23-26], two anthropometric traits (body mass index, height)[27,28], three clinical factors (genetic liabilities to type 2 diabetes, endometriosis, and polycystic ovary syndrome) [29-31], two lifestyle factors (lifetime smoking exposure, circulating 25-hydroxyvitamin D)[32,33], and two molecular risk factors (C-reactive protein, sex hormone-binding globulin) [34,35]. Lifetime smoking exposure is a composite score that captures smoking duration, heaviness, and cessation among both smokers and non-smokers. A step-by-step overview of risk factor inclusion along with a flow-chart of these processes and a list of all risk factors ascertained for inclusion are presented in **Supplementary Materials** and **Supplementary Figure 1**.

Statistical analyses

The use of genetic instruments for potentially modifiable exposures in an MR framework allows for unbiased causal effects of risk factors on disease outcomes to be estimated if: i) the genetic instrument (typically, one or more independent single-nucleotide polymorphisms [SNPs]) is robustly associated with the risk factor of interest; ii) the instrument is not associated with any confounding factor(s) of the association between the risk factor and outcome; and iii) there is no pathway through which an instrument influences an outcome except through the risk factor ("exclusion restriction criterion").

Estimates of the proportion of variance in each risk factor explained by the genetic instruments (R^2) and the strength of the association between the genetic instruments and risk factors (F-statistics) were generated using methods previously described [36]. F-statistics can be used to examine whether results are likely to be influenced by weak instrument bias: i.e., reduced statistical power to reject the null hypothesis when an instrument explains a limited proportion of the variance in a risk factor.

For risk factors with only one SNP as an instrument, the Wald ratio was used to generate effect estimates, and the delta method was used to approximate standard errors [37]; for risk factors with two or three SNPs as instruments, inverse-variance weighted (IVW) fixed effects models were used; and for risk factors with greater than three SNPs, IVW multiplicative random effects models (allowing overdispersion in the model) were used [38]. The combination of multiple SNPs into a multi-allelic IVW model increases the proportion of variance in a risk factor explained by an instrument. Causal estimates from these models represent a weighted average of individual Wald ratios across SNPs using inverse-variance weighted meta-analysis. To account for multiple testing, a Bonferroni correction was used to establish *P*-value thresholds for "strong evidence" ($P<0.0038$) (false positive rate=0.05/13 risk factors) and "suggestive evidence" ($0.0038<P<0.05$) for reported associations.

When using genetic instruments, there is potential for horizontal pleiotropy - when a genetic variant has an effect on two or more traits through independent biological pathways, a violation of the third IV assumption. This was examined by performing three complementary sensitivity analyses, each of which makes different assumptions about the underlying nature of horizontal pleiotropy: i) MR-Egger regression (intercept and slope terms);[39]ii) a weighted median estimator [40] when there were, at minimum, three SNPs in an instrument; and iii) a weighted mode estimator [41] when there were, at minimum, five SNPs in an instrument. Additionally, leave-one-out permutation analyses were performed to examine whether any results were driven by individual SNPs in IVW models. Lastly, Steiger filtering was employed to orient the direction of causal relationships between presumed risk factors and outcomes for some analyses[42]. This method compares the proportion of risk factor and outcome variance explained by SNPs used as instruments to help establish whether SNPs associated with both risk factors and outcomes primarily represent either: 1) a direct association of a SNP on a risk factor which then influences levels of an outcome or 2) a direct association of a SNP on an outcome which then influences levels of a risk factor. Extended descriptions of these sensitivity analyses, along with their assumptions are provided in the **Extended Methods** section.

All statistical analyses were performed using R version 3.3.1.

Results:

Across the 13 risk factors that we examined, F-statistics for their respective genetic instruments ranged from 4 to 423, with 12 of 13 risk factors having a value of $F \geq 24$. These statistics suggest that most analyses were unlikely to suffer from weak instrument bias. For each risk factor, the number of SNPs included in the genetic instrument, along with R^2 and F-

statistics for the instrument, are provided in **Supplementary Table 1**. Complete primary and sensitivity analyses for all risk factors categorized by ovarian cancer histotype are presented in **Supplementary Tables 2-6**.

Reproductive factors

In IVW models, there was suggestive evidence for an effect of earlier age at menarche on risk of overall EOC (OR per year earlier onset: 1.07,95% CI:1.00-1.14;*P*=0.046) and endometrioid carcinoma (OR:1.19,95% CI:1.05-1.36;*P*=0.008) (**Figure 1**). However, there was evidence that horizontal pleiotropy was likely biasing the IVW estimate for EOC. This is because the effect estimate attenuated toward the null when employing MR-Egger regression (OR:1.00,95% CI:0.89-1.13) and a weighted median estimator (OR:1.01,95% CI:0.92-1.10) and moved in a protective direction when using a weighted mode estimator (OR:0.98,95% CI:0.25-3.84). In contrast to EOC, the effect of age at menarche on endometrioid carcinoma was robust to MR-Egger, weighted median, weighted mode estimates, and leave-one-out analyses (**Supplementary Table 2**).

There was suggestive evidence for an effect of later age at natural menopause on risk of endometrioid carcinoma (OR per year later onset:1.09,95% CI:1.02-1.16;*P*=0.007), which was consistent in sensitivity analyses examining horizontal pleiotropy. While there was little evidence of an effect of age at natural menopause on clear cell carcinoma in IVW models (OR:1.05,95% CI:0.96-1.14;*P*=0.29), the association strengthened when employing MR-Egger (OR:1.26,95% CI:1.05-1.52), weighted median (OR:1.11,95% CI:0.99-1.25), and weighted mode estimators (OR:1.16,95% CI:1.02-1.31), suggesting horizontal pleiotropy in the IVW model. There was also suggestive evidence for an effect of genetic liability to twin

births on clear cell carcinoma (OR:1.78,95% CI:1.05-3.03;*P*=0.03) which was robust to sensitivity analyses examining horizontal pleiotropy.

In parity analyses, effect estimates were in a protective direction for five of seven ovarian cancer outcomes but were imprecisely estimated with 95% confidence intervals crossing the null line (**Supplementary Table 2**).

Anthropometric traits

There was strong evidence for an effect of body mass index (BMI) on overall EOC (OR per 1-standard deviation (SD; 4.6 kg/m^2) increase:1.23,95% CI:1.07-1.42;*P*=0.003) (**Figure 2**). Though there was little evidence for horizontal pleiotropy when performing MR-Egger (OR:1.32,95% CI:0.88-1.99), inconsistency of effect estimates across weighted median (OR:1.14,95% CI:0.93-1.40) and weighted mode (OR:1.05,95% CI:0.75-1.51) approaches suggested potential violations of the IV assumptions.

In IVW models, there was suggestive evidence for an effect of BMI on high grade serous carcinoma (OR:1.26,95% CI:1.06-1.50;*P*=0.01), endometrioid carcinoma (OR:1.48,95% CI:1.07-2.06;*P*=0.02), and low malignant potential tumors (OR:1.39,95%CI:1.04-1.85;*P*=0.03) but not on other histotypes. However, there was evidence that horizontal pleiotropy was likely biasing the IVW estimate for high grade serous carcinoma: the effect estimate was attenuated when performing MR-Egger (OR:1.05,95% CI:0.63-1.75) and was inconsistent when employing weighted median (OR:1.17,95% CI:0.91-1.50) and weighted mode (OR:0.95,95% CI:0.53-1.35) estimators. Likewise, there was some inconsistency of effect estimates across sensitivity analyses for low malignant potential tumors, with a modest attenuation of the effect estimate observed when employing a weighted mode estimator (OR:1.17,95%CI:0.55-2.49). In contrast, the effect of BMI on

endometrioid carcinoma was also seen across sensitivity analyses using MR-Egger, weighted median, and weighted mode estimators, and in leave-one-out analyses (**Supplementary Table 3**).

There was strong evidence for an effect of height on clear cell carcinoma (OR per 1-SD (6.3 cm) increase:1.36,95% CI:1.15-1.61;*P*=0.0003), but not on other histotypes. This finding was robust to various sensitivity analyses.

Clinical factors

There was strong evidence for an effect of genetic liability to endometriosis on EOC (per unit log odds higher liability to endometriosis: OR 1.27,95% CI:1.16-1.40;$P=6.94\times10^{-7}$) and clear cell carcinoma (OR:2.69,95% CI:1.88-3.86, $P=7.39\times10^{-8}$) and suggestive evidence for an effect on endometrioid carcinoma (OR:1.37,95% CI:1.10-1.69;*P*=0.004), low malignant potential tumors (OR:1.33,95%CI:1.09-1.63;*P*=0.006), and high grade serous carcinoma (OR:1.17,95% CI:1.04-1.31;*P*=0.007) (**Figure 3**). Findings for overall and clear cell carcinoma were also seen in sensitivity analyses examining horizontal pleiotropy, whereas inconsistent effect estimates for endometrioid carcinoma, low malignant potential tumors, and high grade serous carcinoma across these sensitivity analyses suggested violations of IV assumptions (**Supplementary Table 4**). Analyses employing Steiger filtering provided strong evidence that the causal direction between genetic liability to endometriosis and EOC was from the former to the latter ($P<10^{-10}$), whereas the causal direction could not be clearly established for clear cell carcinoma analyses ($P<0.10$).

There was strong evidence for an inverse effect of genetic liability to polycystic ovary syndrome (PCOS) on endometrioid carcinoma (OR per unit log odds higher liability to PCOS:0.74,95% CI:0.62-0.90;*P*=0.002), which was robust to sensitivity analyses. In contrast,

suggestive evidence for an effect of PCOS with low grade serous carcinoma (OR:1.33,95% CI:1.01-1.74;*P*=0.04) in IVW models was not seen across all sensitivity analyses examining horizontal pleiotropy. There was little evidence of an effect of genetic liability to type 2 diabetes on overall or histotype-specific ovarian cancer.

Lifestyle factors

There was suggestive evidence for an effect of lifetime smoking exposure on EOC (OR per unit increase in smoking score:1.36,95% CI:1.04-1.78,*P*=0.02) (**Figure 4**). In histotype-specific analyses, there was also a suggestive association for an effect of smoking on high grade serous carcinoma (OR:1.44,95% CI:1.05-1.98;*P*=0.02) but little association with other subtypes. The smoking findings for epithelial ovarian cancer and high grade serous carcinoma were robust to horizontal pleiotropy sensitivity analyses (**Supplementary Table 5**). There was no strong or suggestive evidence that circulating 25-hydroxyvitamin D influenced overall or histotype-specific ovarian cancer.

Molecular risk factors

There was suggestive evidence for an inverse effect of C-reactive protein (CRP) on endometrioid carcinoma (OR per unit increase in natural log CRP:0.90,95% CI:0.82-1.00;*P*=0.049) (**Figure 5**). This association was robust to sensitivity analyses using MR-Egger, weighted median, and weighted mode methods in addition to using a restricted CRP instrument (exclusively using 4 SNPs in *CRP*): OR:0.72,95% CI:0.42-1.22;*P*=0.14 (**Supplementary Table 6**). CRP was not clearly associated with other histotypes assessed. There was no strong or suggestive evidence for an effect of sex hormone-binding globulin on ovarian cancer risk.

Discussion

This Mendelian randomization analysis of up to 66,450 women supports causal effects of liability to endometriosis and lifetime smoking exposure in epithelial ovarian cancer risk but found little evidence for causal roles of eleven previously reported risk factors in ovarian carcinogenesis. In histotype-stratified analyses, there was strong or suggestive evidence of effects of ages at menarche and natural menopause, BMI, height, lifetime smoking exposure, CRP and genetic liabilities to twin births and PCOS on ovarian cancer risk. There was little evidence to support causal effects of genetic liability to type 2 diabetes, parity, or circulating levels of 25-hydroxyvitamin D or sex hormone-binding globulin on overall or histotype-specific EOC.

Though historically considered a homogeneous disease with a single cellular origin, epithelial ovarian cancer is now recognized as heterogeneous, consisting of multiple histological subtypes each with its own distinct origins, morphological characteristics, and molecular alterations [18,43-46]. The largely histotype-specific findings in this analysis using genetic variants as proxies to minimize confounding and avoid reverse causation bias thus help to extend these insights further by supporting distinct causal pathways across EOC histotypes.

Some of the histotype-specific findings are consistent with conventional observational studies. For example, in agreement with previous analyses [7-10,47-49], most risk factors did not show clear evidence of association with HGSC. Consistent with some studies, age at natural menopause was most strongly associated with endometrioid carcinoma [8] and height was most strongly associated with clear cell carcinoma [50,51]. The effect of genetic liability to endometriosis on risk of epithelial ovarian cancer is in agreement with two large pooled

observational analyses [9,52], though these studies also reported positive risk relationships with endometrioid and low grade serous carcinoma.

However, some MR estimates were not consistent with those observed in conventional analyses. Most notably, previously reported associations between smoking and mucinous carcinoma [9,53-55] were not corroborated in MR analyses of lifetime smoking exposure. Though estimates from primary and sensitivity analyses all included the null line, inconsistencies in effect estimates across these analyses support pleiotropic biases distorting the causal effect estimate. Though parity has been consistently inversely associated with risk of ovarian cancer in conventional analyses [10,56-60], MR effect estimates suggesting a protective effect of giving birth to more children were imprecise and 95% confidence intervals spanned the null line. Given the few SNPs available to proxy for parity (two independent variants in this analysis), these results likely reflect limited statistical power.

Weaker statistical evidence also suggested an unexpected inverse effect of CRP, a marker of systemic inflammation, on endometrioid carcinoma and positive risk relationships between genetic liability to twin births and clear cell carcinoma. Given recent evidence to suggest a role of infectious agents in ovarian cancer [66, 67], a possible protective effect of CRP on endometrioid carcinoma could speculatively reflect the involvement of CRP in acute immune response (i.e., protection against active bacterial and viral infections). Meanwhile, the effect of genetic liability to twin births on clear cell carcinoma could be mediated by the higher levels of gonadotropins in the fertile years of women with a history of multiple births [54-56].

Overall, few previously reported risk factors showed clear evidence of a causal role in EOC or high grade serous carcinoma, the most common (~70% of cases) and lethal EOC histotype, suggesting that some previously reported associations may have been driven by

residual confounding, misclassification biases, or reverse causation [61]. A notable exception was suggestive evidence that smoking increased odds of HGSC, consistent with some [62,63], but not all [9,53,64,65], observational analyses. A causal effect of genetic liability to endometriosis on EOC corroborates findings from conventional analyses that women with this condition are at elevated risk of subsequent disease[9,66]. This finding also suggests that subclinical manifestations of endometriosis may influence oncogenesis, indicating important avenues for future mechanistic work.

Strengths of this analysis include the use of a systematic approach to collate previously reported risk factors for EOC, the appraisal of the causal role of these risk factors in EOC etiology using a Mendelian randomization framework to reduce confounding and avoid reverse causation bias, the employment of complementary sensitivity analyses to rigorously assess for violations of MR assumptions, and the restriction of datasets utilized to women of primarily or exclusively European descent to minimize confounding through population stratification.

There are several limitations to these analyses. First, though F-statistics generated for most risk factors suggested that results were unlikely to suffer from weak instrument bias, statistical power for some analyses of less common ovarian cancer subtypes (low grade serous, mucinous, and clear cell carcinomas) was likely modest, meaning that the possibility that some results may reflect “false negative” findings cannot be ruled out. Since analyses were performed using summarized genetic association data in aggregate, it was not possible to restrict age at natural menopause analyses exclusively to participants who had undergone menopause. However, given that most ovarian cancer cases occur after menopause and that age-matched controls were used, the inclusion of some pre- or perimenopausal women in these analyses would likely have biased results toward the null (i.e., providing a conservative effect estimate). Additionally, models employed assumed no interaction (e.g., gene-

environment, gene-gene) or effect modification and linear relationships between risk factors and ovarian cancer. Lastly, the use of a MR framework precluded directly examining the causal effects of some ovarian cancer risk factors that do not have robust genetic variants available to serve as proxies (e.g., use of oral contraceptives, hormone replacement therapy).

Though the largely null findings for overall EOC in this analysis can assist in de-prioritizing certain intervention targets for ovarian cancer prevention, they also underscore the challenges in establishing effective primary prevention strategies for this malignancy. To date, beyond risk-reducing surgical interventions, only the oral contraceptive pill has shown compelling evidence that regular use can reduce risk of subsequent disease [59,67,68]. The continued identification of robust genetic variants to proxy other lifestyle and molecular factors previously reported to influence ovarian cancer (e.g., additional sex hormones, gonadotropins, inflammatory markers) will allow for a more refined assessment of the causal influence of these factors in ovarian carcinogenesis [48,69]. Additionally, further work understanding possible mechanisms through which factors that appear to causally influence ovarian cancer in these analyses promote oncogenesis (e.g., genetic liability to endometriosis, C-reactive protein levels) could help to increase scope for prevention opportunities across the life-course. Lastly, for the vast majority of women who develop ovarian cancer with no previous history of smoking and who do not have endometriosis [9,53,70], there is a need to identify novel modifiable risk factors for this condition, as has been advocated elsewhere [71,72].

Conclusions

Of 13 previously reported risk factors examined for association with overall epithelial ovarian cancer, only genetic liability to endometriosis and lifetime smoking exposure showed

evidence compatible with a causal effect on disease risk. When stratified on ovarian cancer histotype, most risk factors showed causal effects on one or more subtypes, underscoring the heterogeneous nature of this disease. While this etiological heterogeneity could have implications for understanding mechanisms of tumour pathology and for studies examining histotype-specific prognosis, given the low incidence of EOC in the general population, prevention strategies targeting factors causally implicated in overall EOC are most likely to confer important population-level reductions in disease incidence. Along with effective clinical management of endometriosis and policies to prevent the initiation of tobacco use and encourage smoking cessation, established prevention strategies like the use of oral contraceptives continue to be important EOC risk-reducing mechanism. The identification of novel modifiable risk factors remains an important priority for the control of epithelial ovarian cancer.

Acknowledgments

The authors' responsibilities were as follows—JY, CLR, RMM, and SJL: conceived the study; JY: planned the analyses, conducted data analysis, and prepared the manuscript; JY, CLR, AL, KM, UM, AG, AW, JZ, GDS, RMM, SJL: critically revised the manuscript; and all authors: read and approved the final manuscript. None of the authors had any conflicts of interest to declare. The authors would like to thank the participants of the individual studies contributing to the Ovarian Cancer Association Consortium (OCAC) for their participation in these studies along with the principal investigators of OCAC for generating the data utilized for this analysis and for making this data available in the public domain.

Funding

This work was supported by a Cancer Research UK programme grant (C18281/A19169) to CLR, SJL, and RMM, including a Cancer Research UK Research PhD studentship (C18281/A20988) to JY. RMM is also supported by the National Institute for Health Research (NIHR) Bristol Biomedical Research Centre. JY, CLR, JZ, GDS, RMM, and SJL are members of the MRC IEU which is supported by the Medical Research Council and the University of Bristol (MC_UU_12013/1-9). UM is supported by the NIHR University College London Hospitals (UCLH) Biomedical Research Centre. JCC is supported by the German Cancer Research Center (DFKZ). WZ is supported by an NIH grant (UM1CA182910). FM is supported by the US National Cancer Institutes (K07-CA80668), the U.S. Army Medical Research and Materiel Command (DAMD17-01-1-0729), and the NIH/National Center for Research Resources/General Clinical Research Center (MO1-RR000056).

References

1. Ferlay J SH, Bray F, et al. *GLOBOCAN 2008 v1.2, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10* Lyon, France: International Agency for Research on Cancer;2010.
2. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. *CA Cancer J Clin.* 2012;62(1):10-29.
3. Henderson BE, Feigelson HS. Hormonal carcinogenesis. *Carcinogenesis.* 2000;21(3):427-433.
4. Noone AM, Howlander N, Krapcho M, et al (eds). *SEER Cancer Statistics Review, 1975-2015.* National Cancer Institute, Bethesda, MD2018.
5. Cress RD, Chen YS, Morris CR, Petersen M, Leiserowitz GS. Characteristics of Long-Term Survivors of Epithelial Ovarian Cancer. *Obstet Gynecol.* 2015;126(3):491-497.
6. Berek JS, Bast RC. Epithelial Ovarian Cancer. In Kufe DW, Pollock RE, Weichselbaum RR, et al., editors. Holland-Frei Cancer Medicine. 6th edition. Hamilton (ON): BC Decker; 2003.
7. Fortner RT, Ose J, Merritt MA, et al. Reproductive and hormone-related risk factors for epithelial ovarian cancer by histologic pathways, invasiveness and histologic subtypes: Results from the EPIC cohort. *Int J Cancer*. 2015;137(5):1196-1208.
8. Gates MA, Rosner BA, Hecht JL, Tworoger SS. Risk factors for epithelial ovarian cancer by histologic subtype. *Am J Epidemiol.* 2010;171(1):45-53.
9. Wentzensen N, Poole EM, Trabert B, et al. Ovarian Cancer Risk Factors by Histologic Subtype: An Analysis From the Ovarian Cancer Cohort Consortium. *J Clin Oncol.* 2016;34(24):2888-2898.

10. Yang HP, Trabert B, Murphy MA, et al. Ovarian cancer risk factors by histologic subtypes in the NIH-AARP Diet and Health Study. *Int J Cancer*. 2012;131(4):938-948.

11. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol*. 2003;32(1):1-22.

12. Yarmolinsky J, Wade KH, Richmond RC, et al. Causal inference in cancer epidemiology: what is the role of Mendelian randomization? *Cancer Epidemiol Biomarkers Prev*. 2018.

13. Amos CI, Dennis J, Wang Z, et al. The OncoArray Consortium: A Network for Understanding the Genetic Architecture of Common Cancers. *Cancer Epidemiol Biomarkers Prev*. 2017;26(1):126-135.

14. Phelan CM, Kuchenbaecker KB, Tyrer JP, et al. Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. *Nat Genet*. 2017;49(5):680-691.

15. Bowtell DD, Bohm S, Ahmed AA, et al. Rethinking ovarian cancer II: reducing mortality from high-grade serous ovarian cancer. *Nat Rev Cancer*. 2015;15(11):668-679.

16. Crane TE, Khulpateea BR, Alberts DS, Basen-Engquist K, Thomson CA. Dietary intake and ovarian cancer risk: a systematic review. *Cancer Epidemiol Biomarkers Prev*. 2014;23(2):255-273.

17. Hunn J, Rodriguez GC. Ovarian cancer: etiology, risk factors, and epidemiology. *Clin Obstet Gynecol*. 2012;55(1):3-23.

18. Karnezis AN, Cho KR, Gilks CB, Pearce CL, Huntsman DG. The disparate origins of ovarian cancers: pathogenesis and prevention strategies. *Nat Rev Cancer*. 2017;17(1):65-74.

19. Trope CG, Kaern J, Davidson B. Borderline ovarian tumours. *Best Pract Res Clin Obstet Gynaecol*. 2012;26(3):325-336.

20. Webb PM, Jordan SJ. Epidemiology of epithelial ovarian cancer. *Best Pract Res Clin Obstet Gynaecol*. 2017;41:3-14.

21. MacArthur J, Bowler E, Cerezo M, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res*. 2017;45(D1):D896-D901.

22. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife*. 2018;7.

23. Barban N, Jansen R, de Vlaming R, et al. Genome-wide analysis identifies 12 loci influencing human reproductive behavior. *Nat Genet.* 2016;48(12):1462-1472.

24. Day FR, Ruth KS, Thompson DJ, et al. Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. *Nat Genet.* 2015;47(11):1294-1303.

25. Day FR, Thompson DJ, Helgason H, et al. Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nat Genet.* 2017;49(6):834-841.

26. Mbarek H, Steinberg S, Nyholt DR, et al. Identification of Common Genetic Variants Influencing Spontaneous Dizygotic Twinning and Female Fertility. *Am J Hum Genet.* 2016;98(5):898-908.

27. Locke AE, Kahali B, Berndt SI, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015;518(7538):197-206.

28. Wood AR, Esko T, Yang J, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet.* 2014;46(11):1173-1186.

29. Albertsen HM, Chettier R, Farrington P, Ward K. Genome-wide association study link novel loci to endometriosis. *PLoS One.* 2013;8(3):e58257.

30. Day F, Karaderi T, Jones MR, et al. Large-Scale Genome-Wide Meta Analysis of Polycystic Ovary Syndrome Suggests Shared Genetic Architecture for Different Diagnosis Criteria. *bioRxiv.* 2018.

31. Morris AP, Voight BF, Teslovich TM, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet.* 2012;44(9):981-990.

32. Jiang X, O'Reilly PF, Aschard H, et al. Genome-wide association study in 79,366 European-ancestry individuals informs the genetic architecture of 25-hydroxyvitamin D levels. *Nat Commun.* 2018;9(1):260.

33. Wootton RE, Richmond RC, Stuijfzand BG, et al. Causal effects of lifetime smoking on risk for depression and schizophrenia: Evidence from a Mendelian randomisation study. *bioRxiv.* 2018;381301.

34. Coviello AD, Haring R, Wellons M, et al. A genome-wide association meta-analysis of circulating sex hormone-binding globulin reveals multiple Loci implicated in sex steroid hormone regulation. *PLoS Genet.* 2012;8(7):e1002805.

35. Dehghan A, Dupuis J, Barbalic M, et al. Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for C-reactive protein levels. *Circulation.* 2011;123(7):731-738.

36. Burgess S, Thompson SG, CRP CHD Genetics Collaboration. Avoiding bias from weak instruments in Mendelian randomization studies. *Int J Epidemiol.* 2011;40(3):755-764.

37. Wald A. The fitting of straight lines if both variables are subject to error. *Ann Math Stat.* 1940;11:284-300.

38. Burgess S, Dudbridge F, Thompson SG. Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods. *Stat Med.* 2016;35(11):1880-1906.

39. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol.* 2015;44(2):512-525.

40. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet Epidemiol.* 2016;40(4):304-314.

41. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol.* 2017;46(6):1985-1998.

42. Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet.* 2017;13(11):e1007081.

43. Li J, Fadare O, Xiang L, Kong B, Zheng W. Ovarian serous carcinoma: recent concepts on its origin and carcinogenesis. *J Hematol Oncol.* 2012;5:8.

44. Kurman RJ. Origin and molecular pathogenesis of ovarian high-grade serous carcinoma. *Ann Oncol.* 2013;24 Suppl 10:x16-21.

45. Seidman JD, Khedmati F. Exploring the histogenesis of ovarian mucinous and transitional cell (Brenner) neoplasms and their relationship with Walthard cell nests: a study of 120 tumors. *Arch Pathol Lab Med*. 2008;132(11):1753-1760.

46. Veras E, Mao TL, Ayhan A, et al. Cystic and adenofibromatous clear cell carcinomas of the ovary: distinctive tumors that differ in their pathogenesis and behavior: a clinicopathologic analysis of 122 cases. *Am J Surg Pathol*. 2009;33(6):844-853.

47. Dixon SC, Nagle CM, Thrift AP, et al. Adult body mass index and risk of ovarian cancer by subtype: a Mendelian randomization study. *Int J Epidemiol.* 2016;45(3):884-895.

48. Ose J, Poole EM, Schock H, et al. Androgens Are Differentially Associated with Ovarian Cancer Subtypes in the Ovarian Cancer Cohort Consortium. *Cancer Res*. 2017;77(14):3951-3960.

49. Ose J, Schock H, Tjonneland A, et al. Inflammatory Markers and Risk of Epithelial Ovarian Cancer by Tumor Subtypes: The EPIC Cohort. *Cancer Epidemiol Biomarkers Prev.* 2015;24(6):951-961.

50. Collaborative Group on Epidemiological Studies of Ovarian Cancer. Ovarian cancer and body size: individual participant meta-analysis including 25,157 women with ovarian cancer from 47 epidemiological studies. *PLoS Med.* 2012;9(4):e1001200.

51. Dixon-Suen SC, Nagle CM, Thrift AP, et al. Adult height is associated with increased risk of ovarian cancer: a Mendelian randomisation study. *Br J Cancer.* 2018.

52. Pearce CL, Templeman C, Rossing MA, et al. Association between endometriosis and risk of histological subtypes of ovarian cancer: a pooled analysis of case-control studies. *Lancet Oncol.* 2012;13(4):385-394.

53. Licaj I, Jacobsen BK, Selmer RM, Maskarinec G, Weiderpass E, Gram IT. Smoking and risk of ovarian cancer by histological subtypes: an analysis among 300 000 Norwegian women. *Br J Cancer*. 2017;116(2):270-276.

54. Jordan SJ, Whiteman DC, Purdie DM, Green AC, Webb PM. Does smoking increase risk of ovarian cancer? A systematic review. *Gynecol Oncol*. 2006;103(3):1122-1129.

55. Modugno F, Ness RB, Cottreau CM. Cigarette smoking and the risk of mucinous and nonmucinous epithelial ovarian cancer. *Epidemiology*. 2002;13(4):467-471.

56. Gaitskell K, Green J, Pirie K, et al. Histological subtypes of ovarian cancer associated with parity and breastfeeding in the prospective Million Women Study. *Int J Cancer*. 2018;142(2):281-289.

57. Riman T, Dickman PW, Nilsson S, et al. Risk factors for invasive epithelial ovarian cancer: results from a Swedish case-control study. *Am J Epidemiol*. 2002;156(4):363-373.

58. Risch HA, Marrett LD, Jain M, Howe GR. Differences in risk factors for epithelial ovarian cancer by histologic type. Results of a case-control study. *Am J Epidemiol*. 1996;144(4):363-372.

59. Tsilidis KK, Allen NE, Key TJ, et al. Oral contraceptive use and reproductive factors and risk of ovarian cancer in the European Prospective Investigation into Cancer and Nutrition. *Br J Cancer*. 2011;105(9):1436-1442.

60. Tung KH, Goodman MT, Wu AH, et al. Reproductive factors and epithelial ovarian cancer risk by histologic type: a multiethnic case-control study. *Am J Epidemiol*. 2003;158(7):629-638.

61. Prat J. New insights into ovarian cancer pathology. *Ann Oncol*. 2012;23 Suppl 10:x111-117.

62. Goodman MT, Tung KH. Active and passive tobacco smoking and the risk of borderline and invasive ovarian cancer (United States). *Cancer Causes Control.* 2003;14(6):569-577.

63. Gram IT, Braaten T, Adami HO, Lund E, Weiderpass E. Cigarette smoking and risk of borderline and invasive epithelial ovarian cancer. *Int J Cancer.* 2008;122(3):647-652.

64. Faber MT, Kjaer SK, Dehlendorff C, et al. Cigarette smoking and risk of ovarian cancer: a pooled analysis of 21 case-control studies. *Cancer Causes Control.* 2013;24(5):989-1004.

65. Collaborative Group on Epidemiological Studies of Ovarian Cancer, Beral V, Gaitskell K, et al. Ovarian cancer and smoking: individual participant meta-analysis including 28,114 women with ovarian cancer from 51 epidemiological studies. *Lancet Oncol.* 2012;13(9):946-956.

66. Wang C, Liang Z, Liu X, Zhang Q, Li S. The Association between Endometriosis, Tubal Ligation, Hysterectomy and Epithelial Ovarian Cancer: Meta-Analyses. *Int J Environ Res Public Health.* 2016;13(11).

67. Bosetti C, Negri E, Trichopoulos D, et al. Long-term effects of oral contraceptives on ovarian cancer risk. *Int J Cancer.* 2002;102(3):262-265.

68. Havrilesky LJ, Moorman PG, Lowery WJ, et al. Oral contraceptive pills as primary prevention for ovarian cancer: a systematic review and meta-analysis. *Obstet Gynecol.* 2013;122(1):139-147.

69. Trabert B, Pinto L, Hartge P, et al. Pre-diagnostic serum levels of inflammation markers and risk of ovarian cancer in the prostate, lung, colorectal and ovarian cancer (PLCO) screening trial. *Gynecol Oncol.* 2014;135(2):297-304.

70. Parkin DM. 2. Tobacco-attributable cancer burden in the UK in 2010. *Br J Cancer*. 2011;105 Suppl 2:S6-S13.

71. Epidemiology Working Group Steering Committee, Ovarian Cancer Association Consortium Members of the EWG SC, Doherty JA, Jensen A, et al. Current Gaps in Ovarian Cancer Epidemiology: The Need for New Population-Based Research. *J Natl Cancer Inst.* 2017;109(10).

72. Tworoger SS, Doherty JA. Epidemiologic paradigms for progress in ovarian cancer research. *Cancer Causes Control.* 2017;28(5):361-364.

Footnote to Supplementary Figure 1

GWAS = genome-wide association study, SNP = single-nucleotide polymorphism, MR = Mendelian randomization, BMI = body mass index, CRP = C-reactive protein, SHBG = sex hormone-binding globulin

Footnote to Figures 1-5

BMI = body mass index, PCOS = polycystic ovary syndrome, 25(OH)D = 25-hydroxyvitamin D, CRP = C-reactive protein, SHBG = sex hormone-binding globulin