Award Number: W81XWH-12-1-0561

TITLE: Development of the Ovarian Cancer Cohort Consortium: Risk Factor Associations by Heterogeneity of Disease

PRINCIPAL INVESTIGATOR: Dr. Shelley Tworoger

CONTRACTING ORGANIZATION: The Brigham and Women's Hospital, Inc.
Boston, MA 02115-6110

REPORT DATE: October 2015

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
October 2015	Annual	30 Sep 2014 – 29 Sep 2015

4. TITLE AND SUBTITLE
Development of the Ovarian Cancer Cohort Consortium: Risk Factor Associations by Heterogeneity of Disease

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-12-1-0561

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Dr. Shelley Tworoger

E-Mail: nhsst@channing.harvard.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Brigham and Women's Hospital, Inc., The
75 Francis Street
Boston, MA 02115-6110

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The objective of this application is to study the etiologic heterogeneity of ovarian cancer in multiple cohorts and to build the infrastructure of the Ovarian Cancer Cohort Consortium (OC3), an international consortium of cohort studies, to address scientific aims important for understanding ovarian cancer risk, early detection, and tumor heterogeneity that are only feasible in a consortium setting. Specifically we will examine associations of risk factors with invasive ovarian cancer, including (but not limited to) age, OCs, tubal ligation, parity, postmenopausal hormone use, family history of ovarian cancer, BMI, height, analgesic use, and lifetime ovulatory cycles, differ by histologic subtype, tumor dominance (as a surrogate for cell of origin), and tumor aggressiveness (tumors fatal within three years vs. all others). Then we will determine if risk prediction models for ovarian cancer can be improved by accounting for differential associations by cancer phenotype. In addition, the proposed efforts will create an infrastructure with a core dataset of important variables for ovarian cancer epidemiology that will be available for future efforts to study ovarian cancer risk, including projects that will use prospectively collected biological specimens. Currently, 23 cohorts have agreed to participate in the OC3. We have executed data use agreements between the Brigham and Women's Hospital (data coordinating center) with all studies. We have received data from 20 cohorts, with 3 cohorts actively preparing data. Data harmonization is complete for the cohorts for which we have received data. Analyses of primary ovarian cancer risk factors (e.g., oral contraceptive use, parity) by histology are complete and a manuscript is drafted and sent to co-authors. Preliminary analyses are on-going for the manuscript examining risk factors by tumor aggressiveness and for the development of a baseline risk prediction model. Several projects have been funded to utilize this resource and additional data collection and variable harmonization are underway.

15. SUBJECT TERMS
Ovarian Cancer, tumor heterogeneity, histology, cell of origin, tumor aggressiveness, risk prediction

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified	98	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Page

INTRODUCTION

The objective of this translational leverage award is to study the etiologic heterogeneity of ovarian cancer in multiple cohorts and to build the infrastructure of the Ovarian Cancer Cohort Consortium (OC3). The OC3 is an international consortium of cohort studies designed to address scientific aims important for understanding ovarian cancer risk, early detection, and tumor heterogeneity. The OC3 is part of the NCI Cohort Consortium, which is an extramural-intramural partnership to address the need for large-scale collaborations and provides the super-structure (but not funding) for managing the OC3. The OC3 currently has 24 participating, on-going cohort studies and we expect there to be over 6,100 invasive ovarian cancer cases among more than 1.4 million women. The goals of the OC3 are to bring together cohorts with ovarian cancer endpoints for pooled projects, build a focused group of ovarian cancer researchers, and develop a comprehensive approach that integrates questionnaire and pathology data with biomarkers, genetics, and tissue. In addition to building the OC3 infrastructure, we propose to evaluate associations of ovarian cancer risk factors by different metrics of tumor heterogeneity. The first specific aim of this application is to examine whether associations of known and putative ovarian cancer risk factors, including (but not limited to) age, oral contraceptive use, tubal ligation, parity, postmenopausal hormone use, family history of ovarian cancer, body mass index, height, analgesic use, and lifetime ovulatory cycles, differ by (a) histologic subtype, (b) tumor dominance (as a surrogate for cell of origin), and (c) tumor aggressiveness (tumors fatal within three years vs. all others). We will use this data to develop ovarian cancer risk prediction models accounting for differential associations by cancer phenotype.

KEYWORDS

Ovarian Cancer, tumor heterogeneity, histology, cell of origin, tumor aggressiveness, risk prediction

OVERALL PROJECT SUMMARY

This grant began on September 30, 2012. Currently, 24 cohorts have agreed to participate in projects addressing the risk factor associations by tumor heterogeneity and to develop an improved risk prediction model for ovarian cancer. The tasks completed in the third year included: (1) invitation of 3 additional cohorts, (2) finalizing data harmonization at the Brigham and Women's Hospital (BWH) data coordinating center (DCC), (4) completing pathologic abstraction for grade and tumor dominance, (5) conducting statistical analyses for our aims, and (6) drafting manuscripts related to the analyses.

A data dictionary and a short questionnaire about the data collection and attributes were sent to all interested cohorts. Only a subset of 10 cohorts have collected pathology reports. Of these, 10 have completed abstraction where possible. One study, has completed coding of tumors that are clearly dominant or non-dominant, but must retrieve records that are in long-term storage from ~120 cases to abstract dimensions for the cases that are uncertain. Below we summarize the cases available for the tumor dominance analysis. For cases with unknown dominance, we have the tumor dimensions for 40% of the cases. We are currently cleaning these data and plan to begin analyses in the no-cost extension.

Table 1. Information on the 10 studies that were able to extract dominance data from pathology reports

Study	N, dom. right	N, dom. left	N, non-dom.	N, Unknown dom.	N, tumor measures
SS	8	11	7	32	20
NYU	33	29	0	67	30
MCCS	32	28	1	113	33
VITAL	28	29	0	106	0
SMC	20	23	0	67	0
WLHS	50	55	2	30	0
NHS	61	68	92	272	128
NHSII	36	57	18	64	46
WHS	25	28	38	113	53
NLCS	120	127	33	204	120
Totals	**413**	**455**	**191**	**1068**	**430**

In total, 24 cohorts from the US, Australia, Europe, and Asia have now agreed to participate. Last year, we invited three additional cohorts to participate in the OC3, including the Million Women's Study (PI: Beral), who declined to participate; the Women's Health Initiative (PI: Anderson), who agreed to a trial participation in one analysis on NSAIDs which may lead to full participation; and the Northern Sweden Health and Disease Study (PI: Lundin and Idahl), who agreed to participate. Also, one study that opted not to participate previously (the Shanghai Women's Health Study) may potentially be willing to participate in a case-cohort design. We are currently negotiating a data use agreement (DUA) with the Swedish group. All other studies have a DUA with the BWH, have provided a letter stating that the IRB does not require a DUA (if sending completely de-identified data), or are BWH-primed cohorts. We received and harmonized data from the Adventist Health Study II, Melbourne Collaborative Cohort Study, and the Swedish Mammography Cohort in the last year and now have data from 23 studies. Details of the participating cohorts including sample sizes are presented in Table 2. Our policies are at our website: https://sites.google.com/a/channing.harvard.edu/oc3/. We are beginning to consider additional questionnaire data types that may be available (Table 2)

Table 2. Details on the OC3 cohorts

Cohort (Acronym)	N[1]	Invasive Cases[2]	Median baseline age	Data available[3]
Adventist Health Study II (AHS2)	46,226	86	54	B
Breast Cancer Detect. Demonstration Proj. (BCDDP)	36,055	145	61	B, FU, D
Breakthrough Generations Study (BGS)	101,881	330	48	B
California Teacher's Study (CTS)	43,782	185	50	B, FU, D
Canadian Study of Diet, Lifestyle, & Health (CSDLH)[4]	39,618	90	58	B, D
Cancer Prevention Study II (CPS2)	65,975	549	62	B, FU, D
Campaign Against Cancer & Heart Disease (CLUEII)	12,393	82	46	B, FU
European Pros. Invest. into Cancer & Nutrition (EPIC)	264,217	704	51	B, D
Iowa Women's Health Study (IWHS)	30,595	268	61	B, FU, D
Melbourne Collab. Cohort Study (MCCS)	23,249	136	55	B, D
Multi-ethnic Cohort Study (MEC)	6,474	75	57	B, FU, D
Netherlands Cohort Study (NCS)[4]	62,573	448	62	B, D
NIH-AARP Diet and Health Study (AARP)	153,084	703	62	B, FU, D
Nurses' Health Study (NHS)	103,298	770	46	B, FU, D
Nurses' Health Study II (NHS2)	111,801	215	35	B, FU, D
NYU Women's Health Study (NYUWHS)	12,431	129	49	B, D
Northern Sweden Health & Disease Study (NSHDS)	43,000	155	NA	B, D
Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO)	60,219	363	62	B, FU, D
Singapore Chinese Health Study (SCHS)	31,945	96	56	B, FU, D
Sister Study (SS)	39,196	39	55	B, FU, D
Swedish Mammography Cohort (SMC)	33,418	39	60	B, FU, D
Vitamins and Lifestyle Study (VITAL)	28,331	130	60	B, D
Women's Health Study (WHS)	33,548	204	53	B, FU, D
Women's Lifestyle & Health Study (WLHS)	49,087	201	40	B, FU
Total	**1,432,396**	**6,142**		

[1]Eligible for inclusion in our analyses, including having a least one ovary and no baseline cancer; [2]There are 491 borderline cases in addition to invasive disease; [3]B=baseline data; FU=Follow-up questionnaires; D=Diet/food frequency questionnaire; [4]Case-cohort design, numbers show full cohort size.

Data harmonization for the key variables is complete for 23 cohorts from which we have received data. Specifically we have cleaned and harmonized the following variables: ovarian cancer diagnosis characteristics (date/age of diagnosis, date of death, type of tumor, morphology, histology, grade), study enrolment and follow-up data (date/age of enrolment, date/age of death, date/age of last follow-up), race, prior cancer

diagnoses, family history of ovarian or breast cancers, menopausal status, postmenopausal hormone use (ever/never, duration, and type), use of oral contraceptives (ever/never, duration), tubal ligation, parity, hysterectomy status, oophorectomy status, age at menarche, age at menopause, smoking, height, body mass index (BMI), BMI at age 18, alcohol intake, endometriosis, other cancer diagnoses, diagnosis of cardiovascular disease, diagnosis of auto-immune disease, diagnosis of diabetes, and NSAIDs. We also have cleaned grade as abstracted from tumor registries or pathology reports; in our initial submission of the histology paper, we were criticized for not examining high and low grade serous tumors separately. To increase power, we abstracted grade from the NHS and NHSII pathology reports (which had not been previously done); in total 17 studies provided grade information. In these studies, among serous tumors, 135 are Grade I, 522 are Grade II, and 1683 are Grade III; 793 have unknown grade. Results of analyses by grade are discussed below.

We have developed SAS macros for conducting analyses in a standardized manner, including a macro to meta-analyze results for a particular exposure across studies, one to conduct a pooled analysis, and macros to assess risk factor association heterogeneity by tumor subtype. We have completed the analysis for examination of ovarian cancer risk factors by histology and a manuscript was submitted to Lancet Oncology. The manuscript was rejected because we did not incorporate grade when examining serous tumors and we did not include endometriosis. Therefore, we added these analyses to the manuscript to preemptively address these criticisms and have submitted to the Journal of Clinical Oncology. The submitted manuscript is in Appendix 1 and the details of the analytic approach are outlined there. Briefly, among over 1.3 million women from 21 studies, 5,510 invasive epithelial ovarian cancers were identified (3331 serous, 592 endometrioid, 334 mucinous, 269 clear cell, 984 other/unknown). Using competing risks Cox proportional hazards regression stratified on study and birth year and adjusted for age, parity, and oral contraceptive use, we assessed associations of 14 ovarian cancer risk factors for all invasive cancers and by histology. Heterogeneity was evaluated by likelihood ratio test. All hormonal/reproductive factors, except breastfeeding and age at menarche, exhibited significant heterogeneity by histology. Higher parity was most strongly associated with endometrioid (RR, per birth=0.79; 95% CI=0.74-0.84) and clear cell (RR=0.67; 95%CI=0.59-0.77) carcinomas (p-het<0.0001). Similarly, age at menopause (positive), endometriosis (positive), and tubal ligation (inverse) were associated with endometrioid and clear cell tumors (p-het<0.004). Family history of breast cancer (p-het=0.008) and body mass index (p-het=0.04) had modest heterogeneity. Smoking was associated with increased risk of mucinous (RR, per 20 pack-years=1.26; 95% CI=1.08-1.46) but a decreased risk of clear cell tumors (RR=0.72; 95% CI=0.55-0.94) (p-het=0.004); height did not have evidence of heterogeneity across types. Among serous tumors, most factors were not differentially associated by grade, although power was limited by the low number of low grade tumors. Endometriosis was significantly associated with low-grade serous tumors (RR: 3.77; 95% CI: 1.24-11.5), but not high-grade serous tumors (RR: 1.11; 95% CI: 0.70-1.74; p-het=0.12). Similarly, more than 5 years of HT use versus never was associated with a 3-fold higher risk of low-grade serous tumors but only a 79% higher risk of high-grade disease, although the p-heterogeneity was not significant (p-het.=0.45). Conversely, family history of ovarian cancer was only significantly associated with high-grade (RR: 1.61, 95% CI: 1.23-2.10) but not low-grade (RR=0.90; 95% CI: 0.22-3.71) serous tumors (p-het.=0.80). Across all exposures, each subtype had unique patterns of risk factor associations. Generally, most risk factors had their strongest association with non-serous cancers. Unsupervised clustering divided the histologic subtypes into two groups based on the similarity of risk factor associations, with serous and mucinous carcinomas in one group and endometrioid and clear cell carcinomas in the other group.

With respect to the rapidly fatal analyses, we had to collect additional mortality data on ovarian cancer cases in 4 studies, who had not provided this data in the initial data transfer. All mortality data have been cleaned and our case definition is as follows: (1) rapidly fatal, death within 3 years of diagnosis, and (2) less aggressive, survived at least three years after diagnosis. In our initial analyses, we required that there be the potential for at least three years of follow-up for all cases to be included, thus excluding some rapidly fatal cases who died <3 before the end of follow-up within a study. However, this reduced power, particularly for analyses in which we further stratified cases by histology (serous vs. endometrioid/clear cell). After discussion with Dr. Peter Kraft, a statistician involved in numerous consortia, we assessed the potential bias of including these cases, and determined that the additional power outweighed any modest bias of a slightly later average diagnosis year for rapidly fatal versus less aggressive cases. Thus we are currently rerunning all analyses to include these cases. Preliminarily, among 4,680 cases with known vital status and the potential for at least 3 years of post-diagnosis

follow-up, 2,257 (49.8%; median survival=1yr) were rapidly fatal and 2,423 (50.2%; median survival=18yr) were less aggressive. Stronger inverse associations were observed for less aggressive than for rapidly fatal disease for tubal ligation (RR, yes vs. no=0.67 vs. 1.07, respectively; p-het=0.001), parity (RR, per child=0.89 vs. 0.93; p-het=0.01), pack years of smoking (RR, per 20 pack-years=0.92 vs. 1.44; p-het=0.01), hormone therapy use (RR, >5 yr vs. never=1.81 vs. 1.48; p-het=0.05), and suggestively for family history of ovarian cancer (RR, yes vs. no=1.73 vs. .29, p-het=0.15). Conversely, women with a BMI>30 vs. >22-25 kg/m^2 were at higher risk of rapidly fatal disease (RR=1.50), but not less aggressive disease (RR=0.99; p-het=0.03). Associations for other risk factors did not differ by aggressiveness. Interestingly, some of the associations for rapidly fatal vs. less aggressive disease differed by histologic subtype. For example, among serous tumors, family history of ovarian cancer was associated with an increased risk of less aggressive disease, but among endometrioid/clear cell tumors, it was associated with an increased risk of rapidly fatal disease. This may be because some of the high-grade endometrioid tumors are truly serous, which have a worse outcome compared to endometrioid tumors. Further, smoking was associated with an increased risk of rapidly fatal serous tumors, but a decreased risk of both rapidly fatal and less aggressive endometrioid/clear cells tumors, consistent with the inverse association of smoking with clear cell disease. Overall our initial analyses support differential associations by tumor aggressiveness for some risk factors. The potentially stronger association of a family history of ovarian cancer with less aggressive disease is supported by reports of better survival in BRCA mutation carriers. The differential association of smoking by tumor aggressiveness, but not that of parity or tubal ligation, may reflect influences of histology. The BMI association with rapidly fatal disease suggests that metabolic dysfunction may play a role in tumor aggressiveness. As noted above, analyses by tumor dominance are on-going.

In addition, progress is being made on the risk prediction model in the OC3 in collaboration with Dr. Ed Iversen at Duke University. Notably, the risk prediction paper using data from the case-control studies in the Ovarian Cancer Association Consortium (OCAC) is under revision at the American Journal of Epidemiology (Appendix 2). The overall AUC in that population including only epidemiologic factors was 0.65 and when adding 17 established low-penetrance genetic alleles was 0.66, suggesting that current genetic risk factors do not substantially increase predictive capability. Interestingly, the AUC was higher for women <50 years (0.71) than women 50 and over (0.62) likely because many risk factors are more strongly associated with endometrioid and clear cell tumors, which are more common in younger women. In the OC3, we are including 8 U.S.-based studies with a minimum set of covariates (e.g., parity, oral contraceptive use) and information on date of diagnosis of ovarian cancer as well as other cancers post-baseline. We excluded the Sister Study because all women have a family history of breast cancer, potentially altering the predictive ability of the model in this higher risk population. Two complete studies have been held out for independent evaluation of the model and the remaining 6 studies were split 80/20 to provide an initial validation set. At this point, the model includes prediction of bilateral salpingo-oophorectomy rates over follow-up (based on data in the Nurses' Health Study and the NHANES dataset), overall mortality, diagnosis of another cancer besides ovarian cancer, and diagnosis of ovarian cancer. Imputation and prediction of risk estimates will use data from both the OCAC and OC3 to increase precision. The following variables have been incorporated into the model: oral contraceptive ever use and duration, family history of breast cancer, family history of ovarian cancer, education (used for imputation), alcohol intake (used for imputation), smoking, endometriosis, age at menarche, tubal ligation, and menopausal status. Other variables (e.g., body mass index, parity, and NSAID use) are being added. Details of the analysis to date are in Appendix 3; this has been run on a 10% sample for error checking purposes. The full dataset will be released by the end of October and we expect the analysis to take about a week to run. This unique collaboration will provide a resource for all future work on ovarian cancer risk prediction, including the incorporation of differential associations by histology.

One of the key goals of the OC3 is to foster collaborations and use of the data nationally and internationally. A list of approved and proposed projects is in Table 3. Importantly, the OC3 is a highly sought after resource. Sixteen projects have been proposed to date from 12 different investigators from 8 institutions; 13 of which have been approved. Three newer proposals (including one from a non-OC3 investigator) will be reviewed at our in-person meeting in Nov. 2015. Through a collaboration with the German Cancer Research Center (DKFZ), we have collected information on blood collection variables as well as existing assay data on several biomarkers from existing nested case-control studies in 7 cohorts. We have cleaned the following additional

variables: case-control status; match set; date and age of blood collection, fasting status, ovarian cancer risk factors at the time of blood draw (not all blood samples were collected at study baseline), time between blood draw and diagnosis, and seven biomarkers (androstenedione, DHEA, DHEAS, testosterone, SHBG, IGF-1, CRP) and the associated assay batch. Preliminary results are shown in Appendix 4. Across the androgens, only testosterone was significantly associated with risk of invasive ovarian cancer overall (RR, doubling=1.12, 95% CI: 1.01-1.23). However, there was significant heterogeneity in the association by histology (p-het.=0.02), with a significant association only for endometrioid (RR, doubling=1.39; 95%CI: 1.02-1.89) and mucinous tumors (RR, doubling=1.29; 95% CI: 1.01-1.66). Interestingly, free testosterone and androstenedione were also significantly positively associated with risk of mucinous carcinoma. Conversely, IGF-1 was significantly inversely associated with risk of invasive ovarian cancer (RR, doubling=0.82; 95%CI: 0.73-0.93), contrary to our initial hypothesis. This association did not differ by histology. Further, in collaboration with Maastricht University, we collected data from 10 studies on peritoneal and fallopian tube cancer cases (some studies do not confirm these tumors and others had already provided this data) to examine risk factor associations by anatomic site. Further, in collaboration with colleagues at the National Cancer Institute, we are evaluating the role of NSAID use with risk of ovarian cancer. This has been a complex data cleaning process using data from 18 studies (included the Women's Health Initiative), with the following variables created for aspirin, non-aspirin NSAIDs (e.g., ibuprofen), and Tylenol: current use at baseline, duration of use, daily dose, and monthly frequency. Also, Dr. Tworoger submitted an R01 to the June 5, 2015 deadline to continue funding for the OC3, focusing on the area of inflammation. The grant was scored in the 27th percentile on its first submission and will be resubmitted March 5, 2016. The aims are in Appendix 5.

Project	Proposed by	Institution	Status
Androgens and risk	Fortner	German Cancer Research Center (DKFZ)	Approved; manuscript being drafted
IGFs and risk	Fortner	DKFZ	Approved; analysis on-going
NSAIDs and risk	Trabert	National Cancer Institute (NCI)	Approved; analysis on-going
Endometriosis and risk	Wentzensen, Trabert	NCI	Approved; incorporated into primary histology paper
CRP/inflammatory factors and risk	Poole, Tworoger	BWH	Approved; submitted R01, Jun 2015 (27th percentile)
Diabetes and risk	Patel, Gapster	American Cancer Society	Approved; incorporated into above R01
OncoArray (GWAS)	Wentzensen, Tworoger	NCI/BWH	Approved; genotyping complete, QC on-going
Risk factors by anatomic sites	Schouten	Univ. of Maastricht	Approved; data cleaning on-going, final DUA negotiations
Proportion of subtype associations explained (methods paper)	Poole, Wentzensen	BWH/NCI	Approved; developing statistical approaches
Hypertension and risk	Huang	BWH	Approved; awaiting new data collection
Exposure-wide association study of high-grade serous tumors	Poole	BWH	Approved; submitting R21, March 2016
Lifecourse adiposity and risk	Fortner, Tworoger	DKFZ/BWH	Approved; awaiting new data collection
Factors associated with long-term survival	Sood	MD Anderson	Approved; DOD grant submitted, Oct. 2015
Telomeres in tumor tissue and survival	Visvanathan	Johns Hopkins	Under review; NIH grant submission, early 2016

Alcohol and risk	Phelan	Moffitt Cancer Center	Under review; submitting R01, March 2016
Lifetime ovulatory cycles and risk	Trabert	NCI	Under review

With respect to the OC3 structure, we continue to have monthly conference calls run by the PI with the Steering Committee. The calls focus on discussing on-going and future collaborations or projects, and vetting preliminary results. Further, given the number of on-going projects, we have a bi-weekly analysis conference call to discuss data cleaning, next steps, and results. This meeting includes Dr. Elizabeth Poole (a junior faculty member working on the project) and the OC3 programmer. The OC3 has had four in-person meetings since the grant started, including at the 2014 Annual NCI Cohort Consortium Meeting. Our next in-person meeting is in November 2015 at the upcoming Cohort Consortium annual meeting. We chose these meeting times because many investigators attend these associated meetings so we have very good attendance. We also have developed a website for the OC3 to communicate our goals, guidelines for participation, and in the future, interesting findings from the study (see https://sites.google.com/a/channing.harvard.edu/oc3/?pli=1).

KEY RESEARCH ACCOMPLISHMENTS

Below is a list of key research accomplishments in the third year of this award.

- Of the 14 established or putative risk factors we examined for ovarian cancer by histologic subtype, 10 risk factors had significant heterogeneity across subtypes.
- Despite having the smallest number of cases, every reproductive/hormonal factor was significantly associated with clear cell tumors, except breastfeeding.
- While endometrioid and clear cell carcinomas had qualitatively similar associations for most risk factors (parity, OC use, age at menopause, tubal ligation, endometriosis, height, family history of ovarian cancer, breastfeeding), they differed in associations related to HT use (which went in opposite directions), family history of breast cancer and BMI (associated with endometrioid only), as well as age at menarche, hysterectomy, and smoking (associated with clear cell only).
- Serous and poorly differentiated carcinomas, the most common and aggressive subtype, had only modest associations for parity, OC use, menopausal HT use, and family history of breast cancer, and stronger associations with family history of ovarian cancer. Further HT use was most strongly associated with low-grade serous tumors. Overall, very few strong risk factors are known for high-grade serous tumors.
- Further, supporting the need to examine associations by histology, androgen levels were only positively associated with endometrioid and mucinous tumors, but not serous or clear cell tumors.
- In unexpected findings, IGF-1 was inversely associated with ovarian cancer risk across all subtypes.
- Most reproductive risk factors were associated preferentially with reducing risk of less aggressive disease, but not rapidly fatal tumors. However, lifestyle factors, such as BMI and smoking, were associated with an increased risk of rapidly fatal tumors, although this association varied by histologic type. This suggests that examining multiple tumor characteristics simultaneously may provide additional etiologic insight.
- Current ovarian cancer risk factors do not have strong predictive capability for identifying specific women at high risk of ovarian cancer, although the AUC is higher for younger women. Given that serous is the most common subtype, but has the least risk factors, it will be critical to identify new risk factors for this type to increase predictive capacity.

CONCLUSION

We are actively developing the OC3 infrastructure by pooling existing cohort data to better elucidate the biology of ovarian cancer. Scientifically, we have or will evaluate whether associations for putative ovarian cancer risk factors differ by tumor subtypes (histology, cell of origin, aggressiveness), as well as develop risk prediction models based on differing risks across subtypes. Further, we are working to develop a “base” risk prediction model that can be used as a comparison for assessing improvement in future work. This will be beneficial to the entire ovarian cancer research community. Importantly in our initial work we observed that

most established or putative ovarian cancer risk factors showed heterogeneity across histologic subtypes and all subtypes had unique patterns of risk factor associations. Endometrioid and clear cell tumors had the strongest associations for many risk factors, and relatively few associations were observed for serous tumors, which are the most common tumor type. This suggests that risk prediction models of ovarian cancer overall will perform worse for serous tumors than for other types. Further, our initial results comparing risk factors for rapidly fatal versus less aggressive disease suggests that this construct adds biologic information beyond that of histology.

Our results support that pre-diagnostic factors may influence ovarian cancer development and aggressiveness and that considering multiple tumor characteristics simultaneously may provide a clearer picture of disease etiology. Ultimately, understanding a woman's risk profile with respect to risk of rapidly fatal versus less aggressive disease at diagnosis may aid in determining the most optimal treatment strategy for long term survival. This has several important implications for etiology and prevention of ovarian cancers. The substantial heterogeneity of individual risk factor associations across ovarian cancer subtypes supports the notion that the subtypes are indeed different diseases and that we may need to consider multiple tumor characterizations to adequately stratify tumors. This underscores the importance of evaluating risk factor and biomarkers associations in consortium settings where there is adequate sample size to provide power to assess associations for the more rare tumor types. The research also suggests that we need to identify new epidemiologic risk factors for serous tumors as the traditional factors are generally most strongly related to endemetrioid and clear cell tumors. Given the higher incidence of serous cancer and its poor survival rates, this is a critical area of future research.

This systematic approach to address ovarian cancer heterogeneity in a large consortial effort will set new standards for evaluating ovarian cancer risk factors and biomarkers and thereby impact understanding of ovarian cancer etiology beyond the work conducted in OC3. Importantly our goal is to continue to expand the data repository of the OC3 by obtaining funding to include dietary factors, updated exposure data from follow-up questionnaires, and biomarker information (both plasma/serum markers and genetics). We also are exploring the possibility of conducting survival analyses. With over 15 projects proposed in the OC3, the development of OC3 infrastructure will have substantial impact on prevention research in the years to com.

PUBLICATIONS, ABSTRACTS, AND PRESENTATIONS

No publications at this time. One manuscript submitted to the Journal of Clinical Oncology and one manuscript under revision at the American Journal of Epidemiology.

Two abstracts were accepted as presentations (presenter is bolded):

1. **Elizabeth M. Poole**, Alan A. Arslan, Lesley M. Butler, James V. Lacey, Jr., I-Min Lee, Alpa V. Patel, Kim Robien, Dale P. Sandler, Leo J. Schouten, V. Wendy Setiawan, Kala Visvanathan, Elisabete Weiderpass, Emily White, Nicolas Wentzensen, Shelley S. Tworoger. Ovarian cancer risk factors by histologic type in the Ovarian Cancer Cohort Consortium (OC3). Presented at the 2014 Annual Meeting of the Society for Epidemiologic Research, June 2014, Seattle, WA.
2. **Shelley S. Tworoger**, Elizabeth M. Poole, Alan A. Arslan, Lesley M. Butler, Victoria Kirsh, James V. Lacey, Jr., I-Min Lee, Alpa V. Patel, Kim Robien, Thomas Rohan, Dale P. Sandler, Leo J. Schouten, V. Wendy Setiawan, Kala Visvanathan, Elisabete Weiderpass, Emily White, Nicolas Wentzensen. Ovarian cancer risk factor associations by tumor aggressiveness in the Ovarian Cancer Cohort Consortium (OC3). Presented at the 10th Biennial Ovarian Cancer Research Symposium sponsored by AACR and the Marsha Rivkin Center for Ovarian Cancer Research, September 2014, Seattle, WA.

Two invited presentations to conference sessions:

1. Elizabeth M. Poole. Ovarian cancer risk factors by histologic type in the Ovarian Cancer Cohort Consortium (OC3). Presented at the Society for Epidemiologic Research Annual Meeting (June 2015). Session: Reproductive Factors and Cancer Risk.

2. Shelley S. Tworoger. Thinking outside the box: New areas in prevention research. Presented at the AACR Advances in Ovarian Cancer Research: Exploiting Vulnerabilities (October 2015). Session: Prevention, Screening, Early Diagnostics, and Epidemiology.

One poster presentation:

1. Nicolas Wentzensen, Elizabeth M. Poole, Alan Arslan, Alpa Patel, V. Wendy Setiawan, Kala Visvanathan, Elisabete Weiderpass, Emily White, Hans-Olov Adami, Louise A. Brinton, Julie Buring, Lesley M. Butler, Tess V. Clendenen, Renee Fortner, Susan M. Gapstur, Mia Gaudet, Patricia Hartge, Judith Hoffman-Bolton, Michael Jones, Vicki Kirsh, Woon-Puay Koh, James V. Lacey, Jr., I-Min Lee, Ulrike Peters, Jenny Poynter, Kim Robien, Thomas Rohan, Dale P. Sandler, Leo J. Schouten, Louise Sjohölm, Anthony Swerdlow, Britton Trabert, Lynne Wilkens, Alicja Wolk, Hannah P. Yang, Anne Zeleniuch-Jacquotte, Shelley S. Tworoger. Ovarian cancer risk factors by histologic subtypes: Evidence for etiologic heterogeneity. AACR Annual Meeting 2015 (Philadelphia, PA).

INVENTIONS, PATENTS, AND LICENCES

None.

REPORTABLE OUTCOMES

The primary reportable outcome is the development of the OC3 database, which contains data on ovarian cancer risk factors and outcomes from 23 cohort studies and by the end of 2015 will contain data from 1 more study. This resource can be used for the analyses proposed in this grant as well as other analyses.

OTHER ACHIEVEMENTS

None.

REFERENCES

None.

APPENDICES

Appendix 1: Submitted manuscript on ovarian cancer risk factor associations by histology, sent to the Journal of Clinical Oncology

Appendix 2: Revised, submitted manuscript outlining the methods for risk prediction modeling in the Ovarian Cancer Association Consortium that are being applied to the OC3

Appendix 3: JAGS output for OC3 risk prediction model algorithm

Appendix 4: Results for androgen and IGF-1 concentrations and risk of ovarian cancer by histology

Appendix 5: Submitted aims for an R01 using the OC3 to examine inflammation and ovarian cancer risk

Appendix 1: Submitted manuscript on ovarian cancer risk factor associations by histology, sent to the Journal of Clinical Oncology

Ovarian cancer risk factors by histologic subtype: An analysis from the Ovarian Cancer Cohort Consortium

Nicolas A. Wentzensen[1*#], Elizabeth M. Poole[2,3*,] Emily White[4], Alan A. Arslan[5], Alpa V. Patel[6], V. Wendy Setiawan[7], Kala Visvanathan[8], Elisabete Weiderpass[9,10,11,12], Hans-Olov Adami[3,11], Amanda Black[1], Leslie Bernstein[13], Louise A. Brinton[1], Julie Buring[3,14], Lesley M. Butler[15], Saioa Chamosa[16], Tess V. Clendenen[5], Laure Dossus[17], Renee Fortner[18], Susan M. Gapstur[6], Mia M. Gaudet[6], Inger T. Gram[9], Patricia Hartge[1], Judith Hoffman-Bolton[8], Annika Idahl[19], Michael Jones[20], Rudolf Kaaks[18], Vicki Kirsh[21], Woon-Puay Koh[22], James V. Lacey,Jr.[13], I-Min Lee[3,14], Eva Lundin[19], Melissa A. Merritt[23], N. Charlotte Onland-Moret [24], Ulrike Peters[4], Jenny N. Poynter[25], Sabina Rinaldi[26], Kim Robien[27], Thomas Rohan[28], Dale P. Sandler[29], Catherine Schairer[1], Leo J. Schouten[30] , Louise K. Sjöholm[11], Sabina Sieri[31], Anthony Swerdlow[20], Anna Tjonneland[32], Britton Trabert[1], Ruth Travis[33], Antonia Trichopoulou[34], Piet A. van den Brandt[30], Lynne Wilkens[35], Alicja Wolk[11], Hannah P. Yang[1], Anne Zeleniuch-Jacquotte[5], Shelley S. Tworoger[2,3]

[1]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health
[2]Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School
[3]Department of Epidemiology, Harvard T.H. Chan School of Public Health
[4]Fred Hutchinson Cancer Research Center, Seattle, WA
[5]New York University School of Medicine
[6]American Cancer Society, Atlanta, GA
[7]University of Southern California, Los Angeles, CA
[8]Johns Hopkins Bloomberg School of Public Health, Baltimore, MD
[9]Department of Community Medicine, Faculty of Health Sciences, University of Tromsø, The Arctic University of Norway
[10]Department of Research, Cancer Registry of Norway
[11]Karolinska Institutet, Stockholm, Sweden
[12]Genetic Epidemiology Group, Folkhälsan Research Center, Helsinki, Finland
[13]City of Hope, Duarte, CA
[14]Division of Preventive Medicine, Brigham and Women's Hospital and Harvard Medical School
[15]Department of Epidemiology, University of Pittsburgh; Cancer Control and Population Sciences, University of Pittsburgh Cancer Institute
[16]Public Health Division of Gipuzkoa, BioDonostia Research Institute, Health Department of Basque Region, San Sebastian, Spain
[17]French Institute of Health and Medical Research
[18]German Cancer Research Center
[19]Umea University
[20]Division of Genetics and Epidemiology and Division of Breast Cancer Research, The Institute of Cancer Research, London, UK
[21]Prevention and Cancer Control, Cancer Care Ontario and Dalla Lana School of Public Health, University of Toronto
[22]Duke-NUS Graduate Medical School Singapore, Singapore
[23]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College of London
[24]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands
[25]Department of Pediatrics, University of Minnesota
[26]International Agency for Research on Cancer, Lyon, France
[27]Department of Exercise and Nutrition Sciences, Milken Institute School of Public Health, George Washington University
[28] Department of Epidemiology and Population Health, Albert Einstein College of Medicine
[29]National Institute of Environmental Health Sciences, National Institutes of Health
[30]GROW-School for Oncology and Developmental Biology - Maastricht University

[31]IRCCS Foundation, National Cancer Institute
[32] Danish Cancer Society Research Center, Denmark
[33]Cancer Epidemiology Unit, Nuffield Department of Population Health University of Oxford Oxford
[34]Hellenic Health Foundation
[35]University of Hawaii Cancer Center

*These authors contributed equally to the manuscript; # Corresponding author

Key words: ovarian cancer, risk factors, subtypes, cohort

To whom correspondence should be addressed:
Nicolas Wentzensen M.D., Ph.D., M.S.
Division of Cancer Epidemiology and Genetics
National Cancer Institute
9609 Medical Center Drive, Room 7-E114
Bethesda, MD 20892-9774
Phone: (240) 276-7303
Fax: (240) 276-7838
wentzenn@mail.nih.gov

Acknowledgements and Grant support:
We would like to thank the participants and staff of the NHS and NHSII for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. The authors assume full responsibility for analyses and interpretation of these data. This work was supported by Department of Defense Ovarian Cancer Research Program grant W81XWH-12-1-0561. The UKBGS thank Breakthrough Breast Cancer and the Institute of Cancer Research for support and funding, and the Study participants, Study staff, and the doctors, nurses and other health care staff and data providers who have contributed to the Study. The ICR acknowledges NHS funding to the NIHR Biomedical Research Centre. K05CA154337 from the National Cancer Institute (NCI) and Office of Dietary Supplements (VITAL). R01 CA39742 (Iowa Women's Health Study). NIH/NCI UM1 CA182876 (SCHS). CA047988, HL043851, HL080467, and HL099355 (Women's Health Study). CA164973 (Multiethnic Cohort). R01 CA77398 and UM1 CA169417 (California Teachers Study). UM1 CA186107, P01 CA87969, UM1 CA176726, R01 CA67262 (Nurses' Health Study, Nurses' Health Study II). The Swedish Women's Lifestyle and Health cohort study was supported by grants of the Swedish Cancer Society and Swedish Research Council. The Swedish Research Council (The Swedish Mammography Cohort). All aspects of the Cancer Prevention Study (CPS)-II were funded by the Intramural Research Program of the American Cancer Society (Atlanta, GA, USA). NCI Intramural Research Program. Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (Z01ES044005; Sister Study). The coordination of EPIC is financially supported by the European Commission (DG-SANCO) and the International Agency for Research on Cancer. The national cohorts are supported by Danish Cancer Society (Denmark); Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l'Education Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM) (France); German Cancer Aid, German Cancer Research Center (DKFZ), Federal Ministry of Education and Research (BMBF); the Hellenic Health Foundation (Greece); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (The Netherlands); Nordic Centre of Excellence programme on Food, Nutrition and Health. (Norway); Health Research Fund (FIS), PI13/00061 to Granada), Regional Governments of Andalucía, Asturias, Basque Country, Murcia (no. 6236) and Navarra, ISCIII RETIC (RD06/0020) (Spain); Swedish Cancer Society, Swedish Scientific Council and County Councils of Skåne and Västerbotten (Sweden); Cancer Research UK (14136 to EPIC-Norfolk; C570/A16491 to EPIC-Oxford), Medical Research Council (1000143 to EPIC-Norfolk) (United Kingdom).

Author contributions:

NW, EMP, SST did the study design, analysis and wrote the first draft of the manuscript. EW, AAA, AVP, WS, KV, and EW were members of the writing group. All authors contributed to data collection, data interpretation and writing of the manuscript.

Abstract
Introduction: Increasing evidence supports that epithelial ovarian cancer is a constellation of diseases with different developmental pathways. We evaluated associations of hormonal, reproductive, and lifestyle factors by histologic subtype in the Ovarian Cancer Cohort Consortium (OC3).
Methods: Among over 1.3 million women from 21 studies, 5,510 invasive epithelial ovarian cancers were identified (3331 serous, 592 endometrioid, 334 mucinous, 269 clear cell, 984 other/unknown). Using competing risks Cox proportional hazards regression stratified on study and birth year and adjusted for age, parity, and oral contraceptive use, we assessed associations for all invasive cancers and by histology. Heterogeneity was evaluated by likelihood ratio test.
Results: All hormonal/reproductive factors, except breastfeeding and age at menarche, exhibited significant heterogeneity by histology. Higher parity was most strongly associated with endometrioid (RR, per birth=0.79; 95% CI=0.74-0.84) and clear cell (RR=0.67; 95%CI=0.59-0.77) carcinomas (p-het<0.0001). Similarly, age at menopause (positive), endometriosis (positive), and tubal ligation (inverse) were associated with endometrioid and clear cell tumors (p-het<0.004). Family history of breast cancer (p-het=0.008) and body mass index (p-het=0.04) had modest heterogeneity. Smoking was associated with increased risk of mucinous (RR, per 20 pack-years=1.26; 95% CI=1.08-1.46) but a decreased risk of clear cell tumors (RR=0.72; 95% CI=0.55-0.94) (p-het=0.004); height did not have evidence of heterogeneity across types.
Discussion: Our results demonstrate heterogeneous associations of risk factors with ovarian cancer subtypes, emphasizing the importance of conducting etiologic studies by ovarian cancer subtypes. Most established risk factors were more strongly associated with non-serous carcinomas, demonstrating challenges for risk prediction of serous cancers, the most fatal subtype.

Introduction

Ovarian cancer is the most lethal gynecologic cancer, with over 152,000 deaths world-wide each year (1). Most ovarian cancers are detected at late stage and have a poor prognosis. Screening for ovarian cancer did not reduce mortality in a large US-based screening trial (2). Understanding the etiologic heterogeneity of ovarian cancer is critical for development of new prevention strategies.

Although multiple carcinogenic mechanisms for ovarian tumorigenesis have been hypothesized, including incessant ovulation, hormonal stimulation, and chronic inflammation (3-6), the etiology of ovarian cancer is not well understood in part due to its heterogeneous nature. Disease subtypes have been categorized by putative precursor lesions, mutations, and histology (7;8). Low-grade serous, mucinous, clear cell, and endometrioid tumors are thought to arise from inclusion cysts or implants in the ovarian surface epithelium and have *K-RAS*, *B-RAF,* or *P-TEN* mutations. High-grade serous tumors, characterized by *TP53* mutations, are thought to arise in the fallopian tube or ovarian epithelium, are more aggressive and have poorer outcomes than other types (7-9). Due to limited power, individual epidemiologic studies usually have considered risk factor associations for all ovarian tumors together. Recently, both individual cohort studies and individual-level meta-analyses of primarily case-control studies have reported differential associations in some ovarian cancer subtypes for menopausal hormone therapy (HT) use, oral contraceptive (OC) use, parity, smoking and body mass index (BMI) (10-16). To establish etiologic models accounting for ovarian cancer heterogeneity, there is a need for a unified prospective evaluation of multiple ovarian cancer risk factors accounting for etiologic heterogeneity. We established the Ovarian Cancer Cohort Consortium (OC3) and evaluated associations of 14 key risk factors with invasive epithelial ovarian cancer risk overall and by histologic subtype based on pooled individual-level data from 5,510 invasive ovarian cancer cases from a combined cohort of over 1.3 million women enrolled in 21 studies.

Methods

Study population

The analysis included women participating in 21 prospective cohort studies from North America, Asia, and Europe (Table 1). Studies were eligible if they had prospective follow-up of ovarian cancer endpoints through

questionnaires, medical records or cancer registries, as well as follow-up for death. Minimal required information included age at study entry, OC use, and parity. All studies obtained institutional approval for cohort maintenance and participation in the OC3. The OC3 Data Coordinating Center and analytic approaches were approved by the institutional review board of the Brigham and Women's Hospital (BWH).

Exposure definitions

Full baseline cohort data (19 studies) or a case-cohort dataset with weights for subcohort members (2 studies) were sent to BWH and were harmonized centrally. Exposures included: parity (ever vs. never, number of births: continuous per 1 birth, and categorical, 1, 2, 3, 4+ births), OC use (ever vs. never, duration of use: continuous, per 5 years of use, and categorical, never, ≤1, >1-≤5, >5-≤10, >10 years), duration of breastfeeding (continuous, per 1 year among parous women), age at menarche (continuous, per 1 year, and categorical, ≤11, 12, 13, 14, ≥15 years), age at natural menopause (postmenopausal women only: continuous, per 5 years, and categorical, ≤40, >40-≤45, >45-≤50, >50-≤55, >55 years), menopausal HT use (ever vs. never, duration of use: continuous, per 1 year, and categorical, never, ≤5, >5 years), tubal ligation (ever vs. never), hysterectomy (ever vs. never), endometriosis (ever vs. never), first degree family history of breast cancer (ever vs. never), first degree family history of ovarian cancer (ever vs. never), BMI (continuous, per 5 kg/m^2, and categorical, <20, 20-<25, 25-<30, 30-<35, ≥35 kg/m^2), height (continuous, per 0.05, and categorical, <1.60, 1.60-<1.65, 1.65-1.70, ≥1.70 m), and smoking (ever vs. never, pack-years: continuous, per 20 pack-years, and categorical, never smoker, ≤10, >10-20, >20-35, >35 pack-years). Studies that did not collect information on a specific risk factor were excluded from the analysis of that factor (Supplemental Table 1), leading to different samples sizes for each variable (Supplemental Table 2).

Outcome definitions

Epithelial ovarian or peritoneal cancer cases were identified either through cancer registries or medical record review (ICD9 codes 183 and 158; ICD10 codes C56). We evaluated associations of risk factors with all invasive epithelial cancers combined (n=5,510). Next, we evaluated associations with the four most common histologic types of invasive epithelial ovarian cancers (n=4,526): serous (including tumors coded as poorly differentiated), endometrioid, mucinous, and clear cell. 984 cases had another histology or were missing histology information and were censored at diagnosis date.

Statistical methods

Women with a history of cancer (other than non-melanoma skin cancer), with bilateral oophorectomy prior to study entry, or with missing age at baseline were excluded from primary analyses. Sensitivity analyses included women with a prior history of cancer. We calculated hazard ratios (HR) and 95% confidence intervals (95% CI) using competing risks Cox proportional hazards regression to evaluate associations between exposures and ovarian cancer endpoints (17). Follow-up time was time between study entry and 1) date of ovarian cancer diagnosis, 2) date of death, or 3) end of follow-up reported by the study, whichever occurred first. In primary analyses, we pooled data from all cohorts, and stratified on year of birth and cohort to account for potential differences in baseline hazards by these factors. Statistical heterogeneity of associations across subtypes was assessed via a likelihood ratio test comparing a model allowing the association for the risk factor of interest to vary by histology versus one not allowing the association to vary (15). We used random effects meta-analysis to combine cohort-specific estimates and to assess between-study heterogeneity. All models were adjusted for age at study entry, number of children, and duration of OC use, unless the exposure of interest was collinear with these factors (e.g., models of ever vs. never parous were not adjusted for number of children). Analysis of hysterectomy was additionally adjusted for HT use. For missing data in covariates (e.g., OC use, parity, and HT use), we filled in missing data with study-specific medians and included a missing indicator in the analysis. Women missing data on a specific exposure of interest were removed from the analysis of that exposure. The Sister Study was excluded from analyses of family history as all participants had a family history of breast or ovarian cancer. To evaluate whether minimally adjusted models (adjusted for age, number of births, and duration of OC use) sufficiently accounted for confounding, we performed a model adjusting for all exposures together. For comparison, we fit our minimally adjusted models in the subset of women with complete information. In 17 studies, grade was available for at least a subset of serous cases. We conducted similar analyses among serous tumors comparing risk factors for low (well-differentiated), moderate (moderately-differentiated), high (poorly-differentiated), and unknown grade. We performed unsupervised

hierarchical clustering of the four subtypes using beta estimates for all exposures analyzed in this study except for duration of breastfeeding (as this factor was not significantly associated with any of the 4 subtypes) using complete linkage and uncentered correlation (Pearson's coefficient). Categories in the cluster analysis were ever vs. never parous, ever vs. never OC use, ever vs. never tubal ligation, ever vs. never endometriosis, age at menarche >15 years vs. <=11 years, age at menopause <40 years vs. 50-55 years, ever vs. never menopausal HT use, ever vs. never hysterectomy, family history of breast cancer (yes vs. no), family history of ovarian cancer (yes vs. no), BMI >35 vs. 20-25, height (per 5cm increase) and ever vs. never smoking. SAS 9.1 was used to conduct the analyses and a p-value of <0.05 was considered statistically significant.

Results

Study population

Among 1,284,090 participants (1,380,779 when considering full cohort size for case-cohort studies), 5,510 invasive epithelial ovarian cancers were identified during follow-up. Cases included in analyses ranged from 1,302 for breastfeeding to 5,510 for OC use (Supplemental Table 2). In total, there were 3,331 (73.6%) serous, 592 (13.1%) endometrioid, 334 (7.4%) mucinous, and 269 (5.9%) clear cell carcinomas. Fifteen of 21 cohorts were based in North America, five in Europe, and one in Asia (Table 1); about half of the cohorts started enrollment in the 1990s. The median age at diagnosis was 66.6 years for serous, 62.0 years for endometrioid, 63.6 years for mucinous, and 60.5 years for clear cell carcinomas.

Associations of hormonal and reproductive factors with ovarian cancer

Most reproductive and hormonal risk factors, except for breastfeeding and hysterectomy, were associated with ovarian cancer risk overall (Table 2). In subtype-specific analyses, a five year increase in duration of OC use was associated with significant 12-16% lower risk of serous, endometrioid, and clear cell carcinomas, but not with mucinous tumors (p-het=0.05). Similarly, OC use longer than 10 years was associated with a 32-50% reduction in risk for serous, endometrioid, and clear cell tumors. Compared to nulliparous women, parous women had a reduced risk of all ovarian cancer subtypes, with significant heterogeneity by subtype (p-het=3.71×10^{-9}). The strongest risk reduction was observed for clear cell (RR: 0.33; 95% CI: 0.25-0.47) carcinomas, while serous cancers had the least risk reduction (RR: 0.79; 95% CI: 0.71-0.88). Similar patterns were observed among parous women for number of children (p-het=3.38×10^{-13}).

A 5-year later menopause was associated with endometrioid and clear cell carcinomas (RR: 1.20; 95% CI: 1.05-1.37 and 1.36; 95% CI: 1.13-1.63, respectively), with a null association for serous (RR: 1.03; 95% CI: 0.98-1.08) and mucinous (RR: 0.90; 95% CI: 0.76-1.06) carcinomas (p-het=0.003). Tubal ligation was only associated with reduced risk of endometrioid (RR: 0.63; 95% CI: 0.43-0.92) and clear cell (RR: 0.36; 95% CI: 0.18-0.70; p-het=0.004) carcinomas, while hysterectomy was inversely associated only with clear cell carcinomas (RR: 0.59; 95% CI: 0.38-0.93; p-het=0.02). Similarly, self-reported endometriosis was strongly associated with endometrioid (RR: 2.47; 95% CI: 1.44-4.23) and clear cell carcinomas (RR: 2.63; 95% CI: 1.37-5.03; p-het=0.03), but was not significantly associated with serous or mucinous tumors. Conversely, a five-year increase in use of menopausal HT was associated with an increased risk of serous (RR: 1.23; 95% CI: 1.19-1.27) and endometrioid (RR: 1.22; 95% CI: 1.12-1.34), but a reduced risk with clear cell (RR: 0.65; 95% CI: 0.47-0.91; p-het=0.00005) carcinomas. There was no significant heterogeneity in associations by histology for duration of breastfeeding or age at menarche, although the latter was significantly inversely associated with clear cell carcinomas.

Among serous tumors, most factors were not differentially associated by grade (Supplemental Table 4). Endometriosis was significantly associated with low-grade serous tumors (RR: 3.77; 95% CI: 1.24-11.5), but not high-grade serous tumors (RR: 1.11; 95% CI: 0.70-1.74; p-het=0.12). Similarly, more than 5 years of HT use versus never was associated with a 3-fold higher risk of low-grade serous tumors but only a 79% higher risk of high-grade disease, although the p-heterogeneity was not significant (p-het.=0.45).

Associations of family history, anthropometric and lifestyle factors with ovarian cancer

Family history of both breast and ovarian cancer and height, but not smoking or BMI were significantly associated with ovarian cancer risk overall (Table 3). A first degree family history of breast or ovarian cancer

was associated with an increased risk of serous tumors (RR, breast: 1.15; 95% CI: 1.03-1.29; RR, ovarian: RR: 1.57; 95% CI: 1.28-1.93), with significant heterogeneity only observed for family history of breast cancer (p-het=0.008). Family history of breast cancer was also associated with endometrioid carcinomas (RR: 1.44; 95% CI: 1.11-1.86). BMI was significantly positively associated with endometrioid carcinomas (RR per 5 kg/m^2: 1.09; 95% CI: 1.00-1.19); but suggestively inversely associated with serous tumors (RR: 0.96; 95% CI: 0.93-1.00; p-het=0.04). Further, each 20 pack-years of smoking was associated with an increased risk of mucinous and a decreased risk of clear cell carcinomas (p-het=0.003). None of these factors were significantly differentially associated by grade among serous tumors (Supplemental Table 4), although family history of ovarian cancer was only significantly associated with high-grade (RR: 1.61, 95% CI: 1.23-2.10) but not low-grade (RR=0.90; 95% CI: 0.22-3.71) serous tumors (p-het.=0.80).

Results for meta-analyses were similar to the pooled analyses (Supplemental Table 3). For example, the RR comparing ever vs. never parous women in the meta-analysis was 0.79 for serous, 0.44 for endometrioid, 0.44 for mucinous and 0.31 for clear cell tumors. We observed little heterogeneity in associations across studies ($p<0.01$ for only 20 of 188 comparisons). Sixteen of associations with between-study heterogeneity were for continuous variables, but the categorical associations did not show heterogeneity. Family history of ovarian cancer showed heterogeneity for all 4 subtypes across studies, but this was likely due to the small number of exposed cases in many of the studies. In sensitivity analyses, inclusion of women with a history of cancer at baseline did not change the results (data not shown). Results were similar when all exposures were included in the model (data not shown).

Patterns of risk factors in histologic subtypes
Each subtype had unique patterns of risk factor associations (Figure 1). The strongest associations for most risk factors were observed for endometrioid and clear cell tumors. Unsupervised clustering divided histologic subtypes into two major groups. Endometrioid and clear cell carcinomas had the most similar risk factor associations (Pearson correlation 0.72). Serous and mucinous cancers were grouped together, but showed more heterogeneity compared to the other two subtypes (Pearson correlation 0.30).

Discussion
In a large pooled analysis of over 1.3 million women, we investigated 14 established or putative risk factors for ovarian cancer by histologic subtype. Ten risk factors had significant heterogeneity across subtypes. Most reproductive and hormonal risk factors had stronger associations with endometrioid and clear cell carcinomas compared to the other types. Serous and poorly differentiated carcinomas, the most common and aggressive subtype, had modest associations for parity, OC use, menopausal HT use, and family history of breast cancer, and stronger associations with family history of ovarian cancer.

Our results are consistent with reports from individual prospective studies within the OC3 (i.e., NHS/NHSII, AARP, and EPIC) (14-16). However, individually these were underpowered to assess subtype-specific associations. Previously, consortia have reported similar subtype-specific associations for individual risk factors, but were largely based on case-control studies (10-13;18;19).

Models of ovarian carcinogenesis have separated epithelial ovarian cancers into major pathways with distinct cells of origin, different carcinogenic pathways and histology with different clinical behavior (7;9). An integrated evaluation of ovarian cancer risk factors by subtypes is important to understand these etiologic pathways on the population level. Each subtype had a qualitatively unique pattern of associations, and serous and mucinous carcinomas were clearly separated from endometrioid and clear cell carcinomas. While endometrioid and clear cell carcinomas had qualitatively similar associations for most risk factors (parity, OC use, age at menopause, tubal ligation, endometriosis, height, family history of ovarian cancer, breastfeeding), they differed in associations related to HT use (which went in opposite directions), family history of breast cancer and BMI (associated with endometrioid only), as well as age at menarche, hysterectomy, and smoking (associated with clear cell only). Despite having the smallest number of cases, every reproductive/hormonal factor was significantly associated with clear cell tumors, except breastfeeding.

Our results further suggest that currently hypothesized, unifying mechanisms, such as incessant ovulation (3), do not apply equally to ovarian cancers. Several variables that determine a woman's lifetime number of ovulations had significant heterogeneity across subtypes. Only parity was similarly associated with all subtypes, suggesting a common biologic effect (20). Notably, mucinous tumors were not associated with any ovulation-related factors except parity, suggesting a more distinct underlying etiology.

Ovarian cancer subtypes share some specific risk factors with other cancer sites. The inverse association between smoking and clear cell ovarian carcinomas is similar to the association of smoking with endometrial cancer (21). Mucinous ovarian cancers share histologic appearance and an association with smoking with colorectal cancers (22). Serous ovarian cancers had weaker associations with most hormonal and reproductive factors compared to non-serous cancers (with the possible exception of OC use), similar to associations observed for hormone receptor negative breast cancers (23). These similarities of risk factor associations across cancers mirror molecular data showing that tumor subtypes from different organs may be more similar to each other on the molecular level compared to other subtypes at the same site (e.g., high-grade serous ovarian cancer and basal-like breast cancer) (24).

While the subtype-specific associations observed in our study strongly corroborate the etiologic heterogeneity of ovarian cancers, a purely histology-based classification of endpoints may have limitations (25). Histologic evaluation is subjective and pathology practice changes over time, which could affect subtype distributions by location and year of diagnosis. For example, we observed the most heterogeneity between studies for mucinous tumors, suggesting that changes in defining mucinous tumors could have led to more variability in associations. However, we did not observe significant differences in subtype proportions across studies or over time (data not shown). We did not observe significant differences in risk factor associations by grade among serous tumors, our results are consistent with a prior study of endometriosis showing an increased risk for low-grade, but not high-grade, tumors. We had relatively few low grade tumors, limiting power. Further, grade reported on pathology reports may not reliable (26); hence we considered moderately differentiated tumors separately. In general, these tumors had similar associations to high-grade tumors. Overall only 5% of serous tumors were low-grade, limiting potential misclassification when considering all serous tumors together (27). Analyses by tumor aggressiveness and tumor dominance have also shown differences in risk factor associations, indicating that there may be important biological heterogeneity beyond histological subtypes (28;29). Further, additional molecular subgroups have been described within high-grade serous ovarian cancers (30;31), but these subtypes have shown only limited heterogeneity in risk factor associations (32).

In summary, we conducted the largest integrated prospective analysis of ovarian cancer risk factors to date. Most risk factors showed heterogeneity across histologic subtypes and each subtype had unique patterns of risk factor associations. Our results have important implications with respect to etiology and prevention of ovarian cancers. Oral contraceptives continue to be an important preventive factor for most types of ovarian cancer. Few other risk factors for ovarian cancer are modifiable and those that are, like smoking and obesity, did not show clear associations with serous carcinomas, the most common and fatal subtype. The substantial heterogeneity of individual risk factor associations across ovarian cancer subtypes supports that subtypes are indeed different diseases and underscores the importance of evaluating risk factors and biomarkers by ovarian cancer subtypes. Our work has implications for the development of risk prediction models, which generally consider ovarian cancer as a whole (33): Due to weaker associations observed for serous carcinomas, prediction of the clinically most important subtype may perform worse than for other types, underscoring the importance of finding better risk markers for serous carcinomas. Evaluation of subtype-specific risk factor and biomarker associations is important for better understanding of ovarian cancer etiology and for targeted development of novel prevention approaches; these analyses require pooling of data for rare subtypes across many studies in consortia.

References

(1) Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 2015 March 1;136(5):E359-E386.

(2) Buys SS, Partridge E, Black A, Johnson CC, Lamerato L, Isaacs C et al. Effect of screening on ovarian cancer mortality: the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Randomized Controlled Trial. JAMA 2011 June 8;305(22):2295-303.

(3) Fathalla MF. Incessant ovulation--a factor in ovarian neoplasia? Lancet 1971 July 17;2(7716):163.

(4) Ness RB, Cottreau C. Possible role of ovarian epithelial inflammation in ovarian cancer. J Natl Cancer Inst 1999 September 1;91(17):1459-67.

(5) Rice MS, Hankinson SE, Tworoger SS. Tubal ligation, hysterectomy, unilateral oophorectomy, and risk of ovarian cancer in the Nurses' Health Studies. Fertil Steril 2014 July;102(1):192-8.

(6) Trabert B, Ness RB, Lo-Ciganic WH, Murphy MA, Goode EL, Poole EM et al. Aspirin, nonaspirin nonsteroidal anti-inflammatory drug, and acetaminophen use and risk of invasive epithelial ovarian cancer: a pooled analysis in the Ovarian Cancer Association Consortium. J Natl Cancer Inst 2014 February;106(2):djt431.

(7) Jarboe EA, Folkins AK, Drapkin R, Ince TA, Agoston ES, Crum CP. Tubal and ovarian pathways to pelvic epithelial cancer: a pathological perspective. Histopathology 2009 November;55(5):619.

(8) Kurman RJ, Shih I. Molecular pathogenesis and extraovarian origin of epithelial ovarian cancer-Shifting the paradigm. Hum Pathol 2011 July;42(7):918-31.

(9) Steffensen KD, Waldstrom M, Grove A, Lund B, Pallisgard N, Jakobsen A. Improved classification of epithelial ovarian cancer: results of 3 danish cohorts. Int J Gynecol Cancer 2011 December;21(9):1592-600.

(10) 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.

(11) Beral V, Doll R, Hermon C, Peto R, Reeves G. Ovarian cancer and oral contraceptives: collaborative reanalysis of data from 45 epidemiological studies including 23,257 women with ovarian cancer and 87,303 controls. Lancet 2008 January 26;371(9609):303-14.

(12) Beral V, Gaitskell K, Hermon C, Moser K, Reeves G, Peto R. Ovarian cancer and smoking: individual participant meta-analysis including 28,114 women with ovarian cancer from 51 epidemiological studies. Lancet Oncol 2012 September;13(9):946-56.

(13) Beral V, Gaitskell K, Hermon C, Moser K, Reeves G, Peto R. Menopausal hormone use and ovarian cancer risk: individual participant meta-analysis of 52 epidemiological studies. Lancet 2015 May 9;385(9980):1835-42.

(14) Fortner RT, Ose J, Merritt MA, Schock H, Tjonneland A, Hansen L 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 February 5.

(15) Gates MA, Rosner BA, Hecht JL, Tworoger SS. Risk factors for epithelial ovarian cancer by histologic subtype. Am J Epidemiol 2010 January 1;171(1):45-53.

(16) Yang HP, Trabert B, Murphy MA, Sherman ME, Sampson JN, Brinton LA et al. Ovarian cancer risk factors by histologic subtypes in the NIH-AARP Diet and Health Study. Int J Cancer 2012 August 15;131(4):938-48.

(17) Lunn M, McNeil D. Applying Cox regression to competing risks. Biometrics 1995 June;51(2):524-32.

(18) Olsen CM, Nagle CM, Whiteman DC, Ness R, Pearce CL, Pike MC et al. Obesity and risk of ovarian cancer subtypes: evidence from the Ovarian Cancer Association Consortium. Endocr Relat Cancer 2013 April;20(2):251-62.

(19) Schouten LJ, Rivera C, Hunter DJ, Spiegelman D, Adami HO, Arslan A et al. Height, body mass index, and ovarian cancer: a pooled analysis of 12 cohort studies. Cancer Epidemiol Biomarkers Prev 2008 April;17(4):902-12.

(20) Adami HO, Lambe M, Persson I, Ekbom A, Hsieh CC, Trichopoulos D. Childbearing and risk of ovarian cancer. Lancet 1995 March 25;345(8952):789.

(21) Setiawan VW, Yang HP, Pike MC, McCann SE, Yu H, Xiang YB et al. Type I and II endometrial cancers: have they different risk factors? J Clin Oncol 2013 July 10;31(20):2607-18.

(22) Newcomb PA, Storer BE, Marcus PM. Cigarette smoking in relation to risk of large bowel cancer in women. Cancer Res 1995 November 1;55(21):4906-9.

(23) Yang XR, Chang-Claude J, Goode EL, Couch FJ, Nevanlinna H, Milne RL et al. Associations of breast cancer risk factors with tumor subtypes: a pooled analysis from the Breast Cancer Association Consortium studies. J Natl Cancer Inst 2011 February 2;103(3):250-63.

(24) Comprehensive molecular portraits of human breast tumours. Nature 2012 October 4;490(7418):61-70.

(25) Kobel M, Bak J, Bertelsen BI, Carpen O, Grove A, Hansen ES et al. Ovarian carcinoma histotype determination is highly reproducible, and is improved through the use of immunohistochemistry. Histopathology 2014 June;64(7):1004-13.

(26) Matsuno RK, Sherman ME, Visvanathan K, Goodman MT, Hernandez BY, Lynch CF et al. Agreement for tumor grade of ovarian carcinoma: analysis of archival tissues from the surveillance, epidemiology, and end results residual tissue repository. Cancer Causes Control 2013 April;24(4):749-57.

(27) Vang R, Shih I, Kurman RJ. Ovarian low-grade and high-grade serous carcinoma: pathogenesis, clinicopathologic and molecular biologic features, and diagnostic problems. Adv Anat Pathol 2009 September;16(5):267-82.

(28) Kotsopoulos J, Terry KL, Poole EM, Rosner B, Murphy MA, Hecht JL et al. Ovarian cancer risk factors by tumor dominance, a surrogate for cell of origin. Int J Cancer 2013 August 1;133(3):730-9.

(29) Poole EM, Merritt MA, Jordan SJ, Yang HP, Hankinson SE, Park Y et al. Hormonal and reproductive risk factors for epithelial ovarian cancer by tumor aggressiveness. Cancer Epidemiol Biomarkers Prev 2013 March;22(3):429-37.

(30) Integrated genomic analyses of ovarian carcinoma. Nature 2011 June 30;474(7353):609-15.

(31) Tothill RW, Tinker AV, George J, Brown R, Fox SB, Lade S et al. Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. Clin Cancer Res 2008 August 15;14(16):5198-208.

(32) Schildkraut JM, Iversen ES, Akushevich L, Whitaker R, Bentley RC, Berchuck A et al. Molecular signatures of epithelial ovarian cancer: analysis of associations with tumor characteristics and epidemiologic risk factors. Cancer Epidemiol Biomarkers Prev 2013 October;22(10):1709-21.

(33) Pfeiffer RM, Park Y, Kreimer AR, Lacey JV, Jr., Pee D, Greenlee RT et al. Risk prediction for breast, endometrial, and ovarian cancer in white women aged 50 y or older: derivation and validation from population-based cohort studies. PLoS Med 2013;10(7):e1001492.

Figure legends:

Figure 1: Unsupervised hierarchical clustering of ovarian cancer histologic subtypes by their associations with hormonal and reproductive risk factors

Unsupervised hierarchical clustering of the four subtypes using the beta estimates using complete linkage, and an uncentered correlation similarity metric. The categories used in the cluster analysis were ever vs. never parous, ever vs. never OC use, ever vs. never tubal ligation, age at menarche >15 years vs. <=11 years, age at menopause <40 years vs. 50-55 years, ever vs. never menopausal HT use, ever vs. never hysterectomy, family history of breast cancer (yes vs. no), family history of ovarian cancer (yes vs. no), BMI >35 vs. 20-25, height (per 5cm increase) and ever vs. never smoking. The color scale shows the range of beta values for each exposure.

Table 1: Characteristics of cohorts participating in the Ovarian Cancer Cohort Consortium

Study name	Study abbreviation	Location	Baseline enrollment period	Baseline cohort size[a]	Median study participant age	Median follow-up (years)	Last year of follow-up	Invasive ovarian cancer cases
NIH-AARP Diet and Health Study	AARP	U.S.	1995-1997	153,084	62	11	2006	703
Breast Cancer Detection Demonstration Project Follow-up Study	BCDDP	U.S.	1987-1989	36,055	61	9	1999	145
Breakthrough Generations Study	BGS	UK	2001-2014	101,881	48	6	2014	75
Canadian Study of Diet, Lifestyle, and Health	CSDLH	Canada	1991-1999	2,745[b]	58	16	2010	90
Campaign against Cancer and Stroke	CLUEII	U.S.	1989	12,393	46	22	2012	82
Cancer Prevention Study II Nutrition Cohort	CPSII-NC	U.S.	1992-1993	65,975	62	15	2009	549
California Teachers Study	CTS	U.S.	1995-1999	43,782	50	15	2010	185
European Prospective Investigation into Cancer and Nutrition Study	EPIC	Europe	1992-2000	264,217	51	13	2010	704
Iowa Women's Health Study	IWHS	U.S.	1986	30,595	61	23	2010	268
Multiethnic/Minority Cohort Study[c]	MEC	U.S.	1993-1998	16,474	57	11	2011	75
Nurses' Health Study 1980[d]	NHS80	U.S.	1980-1982	86,612	46	16	1998	351
Nurses' Health Study 1996[d]	NHS96	U.S.	1996-1998	67,544	62	14	2010	419
Nurses' Health Study II	NHSII	U.S.	1989-1990	111,801	35	20	2011	215
New York University Women's Health Study	NYU	U.S.	1984-1991	12,431	49	24	2012	129
Netherlands Cohort Study on diet and cancer	NLCS	Netherlands	1986	2,757[b]	62	17	2003	448
Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial	PLCO	U.S.	1993-2002	60,219	62	12	2009	363
Singapore Chinese Health Study	SCHS	Singapore	1993-1999	31,945	56	14	2011	96
Sister Study	SS	U.S.	2003-2009	39,196	55	5	2012	39
Swedish Mammography Cohort Study	SMC	Sweden	1997	33,418	60	14	2011	39
VITamins And Lifestyle Cohort	VITAL	U.S.	2000-2002	28,331	60	10	2011	130
Women's Lifestyle and Health	WLHS	Sweden	1991-1992	49,087	40	21	2012	201
Women's Health Study	WHS	U.S.	1993-1996	33,548	53	18	2012	204

[a]After exclusions for baseline cancers and women with bilateral oophorectomy

[b]These cohorts were included as a case-cohort design, reflecting a total cohort population of 39,618 women for the CSDLH and 62,573 women for the NLCS. Appropriate weights for subcohort selection were applied in all analyses.

[c]Including only Caucasian women. [d]The Nurses' Health Study was broken into two study periods (1980-June 1996 and July 1996-2010) because the follow-up was nearly twice as long as any other study. We updated the exposures in 1996 for that follow-up period.

Table 2: Associations[a] of hormonal and reproductive factors with invasive epithelial ovarian cancer overall and by subtypes in the Ovarian Cancer Cohort Consortium

Exposure	All invasive N=5510 RR (95% CI)	Serous N=3331 RR (95% CI)	Endometrioid N=592 RR (95% CI)	Mucinous N=334 RR (95% CI)	Clear cell N=269 RR (95% CI)	p-heterogeneity (between histologic types)[b]
Parity						
Ever/never	0.68 (0.63-0.73)	0.79 (0.71-0.88)	0.48 (0.38-0.59)	0.52 (0.38-0.71)	0.33 (0.25-0.47)	**3.71E-09**
Number of children, per 1 child	0.90 (0.89-0.92)	0.94 (0.91-0.96)	0.79 (0.74-0.84)	0.92 (0.83-1.01)	0.67 (0.59-0.76)	**3.38E-13**
Number of children						
0	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	
1	0.79 (0.71-0.88)	0.83 (0.71-0.96)	0.78 (0.58-1.03)	0.47 (0.29-0.78)	0.65 (0.44-0.96)	
2	0.72 (0.66-0.79)	0.84 (0.75-0.95)	0.49 (0.38-0.63)	0.56 (0.39-0.81)	0.34 (0.24-0.48)	**3.06E-12**
3	0.68 (0.62-0.74)	0.82 (0.72-0.92)	0.41 (0.31-0.54)	0.51 (0.35-0.75)	0.27 (0.17-0.40)	
4+	0.57 (0.52-0.63)	0.69 (0.61-0.79)	0.34 (0.25-0.48)	0.53 (0.35-0.81)	0.14 (0.08-0.26)	
Oral contraceptive use						
Ever/never	0.84 (0.80-0.90)	0.82 (0.76-0.89)	0.89 (0.73-1.08)	1.10 (0.84-1.44)	0.79 (0.60-1.05)	0.21
Duration of use, per 5 year increase	0.86 (0.83-0.90)	0.84 (0.80-0.89)	0.88 (0.79-0.98)	1.05 (0.92-1.21)	0.86 (0.73-1.00)	**0.05**
Duration of use, years						
Never	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	
≤1	1.00 (0.90-1.11)	1.04 (0.92-1.19)	1.03 (0.76-1.39)	0.87 (0.55-1.38)	0.75 (0.46-1.23)	
>1-≤5	0.84 (0.77-0.92)	0.84 (0.75-0.94)	0.81 (0.62-1.05)	0.82 (0.55-1.22)	0.95 (0.66-1.35)	0.32
>5-≤10	0.78 (0.70-0.87)	0.74 (0.65-0.85)	0.90 (0.67-1.20)	0.87 (0.55-1.37)	0.85 (0.55-1.30)	
>10	0.66 (0.57-0.74)	0.62 (0.52-0.73)	0.68 (0.46-0.99)	1.19 (0.74-1.91)	0.50 (0.28-0.89)	
Duration of breastfeeding, per 1 year[c]	0.96 (0.89-1.03)	0.94 (0.86-1.03)	0.85 (0.69-1.05)	0.88 (0.63-1.23)	1.03 (0.80-1.33)	0.64
Age at menarche						
Per 1 year increase	0.98 (0.96-1.00)	0.99 (0.97-1.01)	0.99 (0.94-1.05)	1.02 (0.94-1.10)	0.92 (0.84-1.00)	0.33
Age in years						
≤11	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	
12	0.95 (0.87-1.03)	0.98 (0.87-1.09)	1.02 (0.79-1.32)	1.12 (0.76-1.66)	0.75 (0.51-1.10)	
13	0.94 (0.87-1.02)	1.01 (0.91-1.11)	0.88 (0.69-1.13)	1.07 (0.75-1.53)	0.80 (0.56-1.13)	0.58
14	0.92 (0.83-1.02)	0.97 (0.85-1.10)	0.84 (0.61-1.15)	1.03 (0.65-1.62)	0.80 (0.51-1.27)	
≥15	0.87 (0.78-0.97)	0.91 (0.79-1.04)	1.02 (0.75-1.39)	1.28 (0.84-1.94)	0.56 (0.34-0.94)	
Age at menopause[d]						
Per 5 year increase	1.04 (1.00-1.08)	1.03 (0.98-1.08)	1.20 (1.05-1.37)	0.90 (0.76-1.06)	1.36 (1.13-1.63)	**0.003**
Age in years						
≤40	0.92 (0.79-1.07)	0.90 (0.74-1.09)	0.57 (0.33-1.00)	1.50 (0.84-2.65)	0.15 (0.03-0.74)	0.09
>40-≤45	0.85 (0.74-0.97)	0.93 (0.78-1.10)	0.73 (0.46-1.14)	1.01 (0.54-1.88)	0.43 (0.19-0.97)	

>45-≤50	0.94 (0.87-1.03)	0.97 (0.88-1.08)	0.81 (0.62-1.06)	1.13 (0.77-1.65)	0.89 (0.59-1.35)	
>50-≤55	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	
>55	1.02 (0.88-1.18)	1.01 (0.84-1.22)	1.12 (0.71-1.76)	1.22 (0.64-2.28)	0.96 (0.45-2.03)	
Hormone therapy use[d]						
Ever/never	1.40 (1.31-1.51)	1.48 (1.36-1.61)	1.72 (1.37-2.14)	1.02 (0.74-1.40)	0.90 (0.62-1.30)	**0.004**
Duration of use, per 5 year increase	1.21 (1.17-1.24)	1.23 (1.19-1.27)	1.22 (1.12-1.34)	1.11 (0.96-1.30)	0.65 (0.47-0.91)	**0.00005**
Duration of use, years						
Never	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	
≤5 years	1.18 (1.08-1.30)	1.26 (1.12-1.41)	1.54 (1.16-2.06)	1.08 (0.72-1.61)	0.93 (0.60-1.46)	**0.0002**
>5 years	1.63 (1.49-1.79)	1.83 (1.64-2.04)	1.77 (1.31-2.40)	1.14 (0.71-1.80)	0.46 (0.23-0.92)	
Tubal ligation, ever/never	0.86 (0.76-0.97)	0.95 (0.82-1.10)	0.63 (0.43-0.92)	1.07 (0.63-1.82)	0.36 (0.18-0.70)	**0.004**
Hysterectomy[e]**,** ever/never	1.04 (0.96-1.12)	1.09 (0.99-1.20)	0.98 (0.77-1.25)	0.82 (0.57-1.17)	0.59 (0.38-0.93)	**0.02**
Endometriosis, ever/never	1.35 (1.07-71)	1.08 (0.77-1.52)	2.47 (1.44-4.23)	1.69 (0.60-4.71)	2.63 (1.37-5.03)	**0.03**

[a]Stratified on birth year and cohort, and adjusted for age at study entry, parity, and duration of oral contraceptive use (except when parity or oral contraceptive use was the primary exposure of interest and then we adjusted only for the other risk factor) using pooled analyses of all cohorts combined.
[b]Assessed using a likelihood ratio test comparing a Cox proportional hazards competing risks model allowing the association to vary by histologic subtype to a model forcing the association to be the same across subtypes.
[c]Parous women only.
[d]Postmenopausal women only.
[e]Additionally adjusted for duration of hormone therapy use.

Table 3: Associations[a] of family history, demographic and lifestyle factors with invasive epithelial ovarian cancer overall and by subtypes in the Ovarian Cancer Cohort Consortium

Exposure	All invasive N=5510 RR (95% CI)	Serous N=3331 RR (95% CI)	Endometrioid N=592 RR (95% CI)	Mucinous N=334 RR (95% CI)	Clear cell N=269 RR (95% CI)	p-diff (between histologic types)[b]
First degree family history of breast cancer, ever/never	1.13 (1.03-1.23)	1.15 (1.03-1.29)	1.44 (1.11-1.86)	0.77 (0.48-1.22)	0.63 (0.35-1.09)	**0.008**
First degree family history of ovarian cancer, ever/never	1.46 (1.24-1.73)	1.57 (1.28-1.93)	0.98 (0.52-1.84)	1.34 (0.59-3.03)	0.96 (0.36-2.58)	0.39
Body mass index						
Per 5 kg/m^2	1.01 (0.98-1.04)	0.96 (0.93-1.00)	1.09 (1.00-1.19)	1.05 (0.94-1.19)	1.02 (0.91-1.15)	**0.04**
In kg/m^2						
<20	1.04 (0.93-1.16)	1.09 (0.94-1.26)	0.84 (0.58-1.20)	1.47 (0.94-2.27)	0.93 (0.57-1.52)	
20-<25	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	
25-<30	0.95 (0.89-1.02)	0.90 (0.82-0.98)	0.98 (0.80-1.21)	1.60 (1.21-2.11)	1.18 (0.88-1.60)	**0.01**
30-<35	0.97 (0.88-1.07)	0.91 (0.80-1.03)	1.13 (0.85-1.50)	1.22 (0.78-1.90)	0.87 (0.54-1.41)	
≥35	1.10 (0.97-1.25)	0.98 (0.83-1.15)	1.35 (0.94-1.94)	1.09 (0.57-2.11)	1.17 (0.66-2.09)	
Height						
Per 0.5m	1.06 (1.03-1.08)	1.05 (1.02-1.08)	1.05 (0.98-1.12)	1.03 (0.94-1.13)	1.07 (0.96-1.19)	0.96
In meters						
<1.60	0.89 (0.82-0.96)	0.88 (0.79-0.97)	1.01 (0.80-1.28)	0.80 (0.58-1.11)	0.91 (0.64-1.29)	
1.60-<1.65	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	0.50
1.65-<1.70	1.03 (0.96-1.11)	1.05 (0.96-1.16)	0.93 (0.73-1.19)	0.87 (0.63-1.21)	0.91 (0.64-1.30)	
≥1.70	1.11 (1.02-1.21)	1.04 (0.94-1.16)	1.22 (0.96-1.56)	1.02 (0.72-1.45)	1.22 (0.86-1.74)	
Smoking						
Ever/never	1.01 (0.95-1.07)	0.99 (0.92-1.07)	0.98 (0.82-1.18)	1.43 (1.11-1.84)	1.01 (0.78-1.30)	**0.05**
Per 20 pack-years	1.00 (0.96-1.04)	1.02 (0.97-1.07)	0.95 (0.82-1.10)	1.26 (1.08-1.46)	0.72 (0.55-0.94)	**0.003**
In pack-years						
Never	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	
≤10	1.07 (0.97-1.19)	1.04 (0.91-1.18)	1.07 (0.80-1.44)	1.34 (0.86-2.08)	0.92 (0.60-1.41)	
>10-20	1.05 (0.93-1.20)	1.06 (0.90-1.24)	0.73 (0.46-1.13)	1.70 (1.00-2.89)	1.04 (0.61-1.77)	0.09
>20-35	1.01 (0.89-1.15)	1.06 (0.90-1.24)	0.94 (0.64-1.39)	1.34 (0.78-2.31)	0.46 (0.21-1.00)	
>35	1.03 (0.91-1.17)	1.10 (0.95-1.28)	0.98 (0.65-1.48)	1.84 (1.11-3.05)	0.46 (0.20-1.04)	

[a]Stratified on birth year and cohort, and adjusted for age at study entry, parity, and duration of oral contraceptive use (except when parity or oral contraceptive use was the primary exposure of interest and then we adjusted only for the other risk factor) using a pooled analysis of all cohorts combined.

[b]Assessed using a likelihood ratio test comparing a Cox proportional hazards competing risks model allowing the association to vary by histologic subtype to a model forcing the association to be the same across subtypes.

Supplemental Table 1. Studies[a] in the Ovarian Cancer Cohort Consortium contributing to each exposure analysis

Variable	Studies
Ever/never parous:	AARP, BCDDP, BGS, CLUEII, CPSII-NC, CSDLH, CTS, EPIC, IWHS, MEC, NHS, NHSII, NLCS, NYU, PLCO, SCHS, SMC, SS, VITAL, WHS, WLHS
Number of children (continuous or categorical):	AARP, BCDDP, BGS, CLUEII, CPSII-NC, CSDLH, CTS, EPIC, IWHS, MEC, NHS, NHSII, NLCS, NYU, PLCO, SMC, SS, VITAL, WHS, WLHS
Ever/never OC use:	AARP, BCDDP, BGS, CLUEII, CPSII-NC, CSDLH, CTS, EPIC, IWHS, MEC, NHS, NHSII, NLCS, NYU, PLCO, SCHS, SMC, SS, VITAL, WHS, WLHS
Duration of OC use (continuous or categorical):	AARP, BCDDP, BGS, CLUEII, CPSII-NC, CSDLH, CTS, EPIC, IWHS, MEC, NHS, NHSII, NLCS, NYU, PLCO, SCHS, SMC, SS, VITAL, WHS, WLHS
Duration of breastfeeding (continuous):	BGS, CTS, EPIC, NHS, NHSII, SS, WLHS
Age at menarche (continuous or categorical):	AARP, BCDDP, BGS, CLUEII, CPSII-NC, CSDLH, CTS, EPIC, IWHS, MEC, NHS, NHSII, NLCS, NYU, PLCO, SCHS, SMC, SS, VITAL, WHS, WLHS
Age at menopause (continuous and categorical):	AARP, BCDDP, BGS, CLUEII, CPSII-NC, CSDLH, CTS, EPIC, IWHS, MEC, NHS, NLCS, NYU, PLCO, SCHS, SMC, SS, VITAL, WHS
Ever use of HT	AARP, BCDDP, BGS, CLUEII, CPSII-NC, CSDLH, CTS, EPIC, IWHS, MEC, NHS, NLCS, NYU, PLCO, SCHS, SMC, SS, VITAL, WHS, WLHS
Duration of HT use (continuous and categorical):	AARP, BCDDP, BGS, CPSII-NC, CSDLH, EPIC, IWHS, MEC, NHS, NLCS, NYU, PLCO, SCHS, SMC, SS, VITAL, WHS
Tubal ligation:	CPSII-NC, CTS, EPIC, MEC, NHS, NHSII, NLCS, NYU, PLCO, SMC, SS, VITAL, WHS
Hysterectomy:	AARP, BCDDP, BGS, CLUEII, CPSII-NC, CSDLH, EPIC, IWHS, MEC, NHS, NHSII, NLCS, NYU, PLCO, SCHS, SMC, SS, VITAL, WHS
Endometriosis:	BGS, CTS, IWHS, NHSII, PLCO, SS
Family history of breast cancer:	AARP, BCDDP, BGS, CLUEII, CPSII-NC, CSDLH, CTS, EPIC, IWHS, MEC, NHS, NHSII, NLCS, NYU, PLCO, SCHS, SMC, VITAL, WHS
Family history of ovarian cancer:	AARP, BCDDP, BGS, CLUEII, CPSII-NC, CTS, IWHS, MEC, NHS, NHSII, NLCS, PLCO, SCHS, SS, VITAL, WHS
BMI (continuous and categorical):	AARP, BCDDP, BGS, CLUE, CPSII-NC, CSDLH, CTS, EPIC, IWHS, MEC, NHS, NHSII, NLCS, NYU, PLCO, SCHS, SMC, SS, VITAL, WHS, WLHS
Height (continuous and categorical):	AARP, BCDDP, BGS, CLUE, CPSII-NC, CSDLH, CTS, EPIC, IWHS, MEC, NHS, NHSII, NLCS, NYU, PLCO, SCHS, SMC, SS, VITAL, WHS, WLHS
Ever/never smoker:	AARP, BCDDP, BGS, CLUEII, CPSII-NC, CSDLH, CTS, EPIC, IWHS, MEC, NHS, NHSII, NLCS, NYU, PLCO, SCHS, SMC, SS, VITAL, WHS, WLHS
Pack-years of smoking (continuous and categorical):	BCDDP, BGS, CPSII-NC, CSDLH, IWHS, MEC, NHS, NHSII, NLCS, NYU, PLCO, SCHS, SMC, SS, VITAL, WHS

[a]Study abbreviations can be found in Table 1

Supplemental Table 2. Number of invasive epithelial ovarian cancer cases overall and by histologic subtype for each exposure

Case numbers for each exposure	Serous	Endometrioid	Mucinous	Clear cell	All Invasive
Parity					
Ever/never	3248	582	321	254	5352
Number of children (continuous or categorical)	3208	568	303	238	5273
Oral contraceptive use					
Ever/never	3331	592	334	269	5510
Duration of use (continuous or categorical)	3198	567	314	259	5271
Duration of breastfeeding	827	157	69	64	1302
Age at menarche (continuous or categorical)	3283	587	329	267	5417
Age at menopause (postmenopausal only; continuous or categorical)	2124	337	208	132	3449
HT use (postmenopausal only)					
Ever/never	2557	392	228	149	4243
Duration of use (continuous or categorical)	2335	333	217	136	3726
Tubal ligation	2337	420	214	193	3848
Hysterectomy	3287	582	326	258	5412
Endometriosis	806	146	70	82	1391
First degree family history of breast cancer	3219	571	319	258	5309
First degree family history of ovarian cancer	2649	462	242	206	4347
Body mass index (continuous or categorical)	3186	563	321	262	5281
Height (continuous or categorical)	3227	577	324	267	5357
Smoking					
Ever/never	3284	589	330	268	5440
Pack-years(continuous or categorical)	2158	379	217	187	4520

Supplemental Table 3. Associations[a] of risk factors with ovarian cancer subtypes based on meta-analysis pooling the results of individual studies in the Ovarian Cancer Cohort Consortium

Exposure	Serous	Endometrioid	Mucinous	Clear cell
Parity				
Ever/never	0.79 (0.71-0.87)	0.44 (0.34-0.55)	0.44 (0.31-0.63)	0.31 (0.23-0.42)
Number of children, per 1 child	0.93 (0.91-0.96)	0.81 (0.71-0.92)[b]	0.86 (0.75-0.97)[b]	0.59 (0.49-0.72)[b]
Number of children				
0	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)
1	0.85 (0.72-1.00)	0.78 (0.57-1.07)	0.63 (0.40-0.99)	0.57 (0.35-0.92)
2	0.84 (0.75-0.95)	0.49 (0.39-0.63)	0.55 (0.40-0.78)	0.38 (0.24-0.59)
3	0.81 (0.71-0.91)	0.44 (0.34-0.57)	0.48 (0.30-0.77)	0.30 (0.18-0.51)
4+	0.69 (0.60-0.80)	0.34 (0.23-0.48)	0.55 (0.38-0.80)	0.35 (0.14-0.85)
Oral contraceptive use				
Ever/never	0.83 (0.76-0.90)	0.88 (0.72-1.07)	1.11 (0.85-1.46)	0.76 (0.54-1.06)
Duration of use, per 5 year increase	0.85 (0.79-0.91)	0.90 (0.77-1.04)	1.23 (0.91-1.65)[b]	0.96 (0.82-1.11)
Duration of use, years				
Never	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)
≤1	1.06 (0.92-1.22)	1.17 (0.84-1.63)	1.09 (0.69-1.74)	1.36 (0.79-2.35)
>1-≤5	0.88 (0.78-0.99)	0.92 (0.70-1.21)	1.12 (0.70-1.78)	1.39 (0.83-2.33)
>5-≤10	0.81 (0.69-0.94)	0.95 (0.70-1.28)	1.36 (0.88-2.11)	1.11 (0.67-1.83)
>10	0.67 (0.56-0.81)	0.78 (0.46-1.31)	1.56 (0.94-2.59)	0.75 (0.32-1.74)
Duration of breastfeeding, per 1 year[c]	1.01 (0.87-1.18)[b]	0.93 (0.78-1.11)	0.94 (0.68-1.31)	1.13 (0.93-1.36)
Age at menarche				
Per 1 year increase	0.99 (0.96-1.02)	1.02 (0.97-1.08)	1.08 (0.96-1.22)[b]	0.96 (0.91-1.02)
Age in years				
≤11	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)
12	0.96 (0.82-1.14)	0.95 (0.72-1.25)	1.16 (0.73-1.84)	0.77 (0.49-1.20)
13	1.02 (0.92-1.13)	0.94 (0.71-1.24)	1.07 (0.73-1.57)	0.83 (0.44-1.59)
14	0.98 (0.85-1.13)	0.84 (0.59-1.19)	1.07 (0.63-1.80)	0.77 (0.45-1.32)
≥15	0.92 (0.77-1.10)	1.00 (0.70-1.42)	1.50 (0.90-2.48)	0.75 (0.39-1.42)
Age at menopause				
Per 5 year increase	1.04 (0.99-1.09)	1.39 (1.02-1.89)[b]	1.07 (0.78-1.47)[b]	2.06 (1.38-3.08)[b]
Age in years				
≤40	0.99 (0.81-1.21)	0.81 (0.46-1.40)	2.00 (0. 67-5.29)	0.64 (0.14-2.89)
>40-≤45	0.95 (0.79-1.13)	0.96 (0.64-1.44)	1.23 (0.74-2.03)	1.06 (0.35-3.22)
>45-≤50	0.97 (0.87-1.08)	0.79 (0.59-1.05)	1.18 (0.85-1.63)	1.02 (0.65-1.59)
>50-≤55	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)
>55	1.06 (0.88-1.28)	1.17 (0.76-1.80)	2.03 (0.96-4.27)	2.00 (0.91-4.38)

HT use[d]				
Ever/never	1.47 (1.34-1.61)	1.84 (1.44-2.36)	1.08 (0.77-1.50)	0.94 (0.57-1.55)
Duration of use, per 5 year increase	1.24 (1.18-1.31)	1.30 (1.13-1.49)	1.21 (0.93-1.58)	0.49 (0.28-0.84)[b]
Duration of use, years				
Never	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)
≤5	1.27 (1.13-1.42)	1.86 (1.31-2.64)	1.23 (0.81-1.88)	1.08 (0.61-1.90)
>5	1.85 (1.65-2.07)	2.22 (1.46-3.38)	1.53 (0.93-2.53)	0.95 (0.80-1.13)
Tubal ligation, ever/never	0.98 (0.82-1.17)	0.80 (0.53-1.19)	1.43 (0.80-2.56)	0.63 (0.27-1.46)
Hysterectomy, ever/never[e]	1.04 (0.92-1.17)	1.20 (0.71-2.02)[b]	0.87 (0.60-1.27)	0.87 (0.53-1.44)
Endometriosis, yes/no	1.14 (0.81-1.61)	2.84 (1.56-5.18)	5.09 (1.54-16.9)	3.44 (1.52-7.79)
First degree family history of breast cancer, yes/no	1.21 (1.04-1.41)	1.54 (1.19-2.00)	1.13 (0.70-1.81)	1.04 (0.59-1.84)
First degree family history of ovarian cancer, yes/no	0.97 (0.35-2.71)[b]	0.26 (0.00-16.8)[b]	0.01 (0.00-6.61)[b]	0.04 (0.00-8.57)[b]
Body mass index				
Per 5 kg/m^2	0.97 (0.93-1.01)	1.00 (0.87-1.15)[b]	1.06 (0.89-1.26)[b]	0.95 (0.82-1.10)[b]
In kg/m^2				
<20	1.11 (0.97-1.29)	1.14 (0.78-1.65)	1.77 (1.17-2.67)	1.34 (0.81-2.23)
20-<25	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)
25-<30	0.92 (0.82-1.03)	1.02 (0.82-1.26)	1.68 (1.26-2.25)	1.30 (0.95-1.77)
30-<35	0.93 (0.82-1.05)	1.35 (1.00-1.81)	1.95 (1.23-3.10)	1.59 (0.93-2.73)
≥35	1.05 (0.82-1.35)	1.75 (1.21-2.54)	1.96 (0.96-4.03)	2.08 (1.07-4.06)
Height				
Per 0.5m	1.05 (1.02-1.08)	1.04 (0.97-1.12)	1.07 (0.95-1.20)[b]	1.12 (1.07-1.17)
In meters				
<1.60	0.88 (0.80-0.98)	1.03 (0.82-1.30)	0.91 (0.64-1.29)	0.96 (0.66-1.39)
1.60-<1.65	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)
1.65-<1.70	1.06 (0.94-1.20)	1.01 (0.80-1.26)	0.95 (0.70-1.27)	0.92 (0.62-1.35)
≥1.70	1.04 (0.93-1.16)	1.20 (0.93-1.55)	1.02 (0.75-1.38)	1.19 (0.81-1.75)
Smoking				
Ever/never	1.02 (0.91-1.13)	1.02 (0.85-1.22)	1.37 (1.06-1.78)	0.95 (0.71-1.27)
Continuous pack-years, per 20 pack-years	1.05 (0.98-1.12)	1.00 (0.87-1.16)	0.77 (0.48-1.22)[b]	0.62 (0.34-1.13)[b]
Categorical pack-years				
Never	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)
≤10	1.10 (0.96-1.25)	1.25 (0.93-1.69)	1.46 (0.93-2.30)	0.99 (0.61-1.61)
>10-20	1.08 (0.91-1.28)	0.87 (0.55-1.39)	1.21 (0.79-1.84)	1.27 (0.67-2.41)
>20-35	1.15 (0.98-1.35)	1.20 (0.79-1.81)	1.43 (0.88-2.32)	0.81 (0.34-1.95)
>35	1.11 (0.93-1.32)	1.18 (0.76-1.83)	1.58 (0.83-3.02)	0.98 (0.40-2.40)

[a]Stratified on birth year, and adjusted for age at study entry, parity, and duration of oral contraceptive use (except when parity or oral contraceptive use was the primary exposure of interest and then we adjusted only for the other risk factor).
[b]Meta-analysis p-heterogeneity across studies <0.01 using the q-statistic from a random-effects meta-analysis.
[c]Parous women only.
[d]Postmenopausal women only.
[e]Additionally adjusted for duration of hormone therapy use.

Supplemental Table 4. Associations[a] of risk factors with among serous ovarian carcinomas by grade in the Ovarian Cancer Cohort Consortium

Exposure	Well-differentiated[b]	Moderately-differentiated	Poorly-differentiated	Unknown grade	p-het.[c]
Parity					
Ever/never	0.72 (0.43-1.21)	0.78 (0.60-1.02)	0.82 (0.71-0.96)	0.83 (0.67-1.04)	0.12
Number of children, per 1 child	0.90 (0.80-1.02)	0.89 (0.84-0.95)	0.93 (0.91-0.96)	0.96 (0. 19-1.00)	0.33
Number of children					
0	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	
1	0.71 (0.33-1.52)	0.90 (0.62-1.30)	0.86 (0.70-1.07)	0.87 (0.64-1.18)	
2	0.79 (0.44-1.42)	0.86 (0.64-1.16)	0.87 (0.74-1.03)	0.84 (0.65-1.07)	0.65
3	0.82 (0.46-1.47)	0.71 (0.52-0.99)	0.87 (0.73-1.03)	0.84 (0.65-1.07)	
4+	0.45 (0.22-0.94)	0.66 (0.47-0.93)	0.67 (0.55-0.80)	0.84 (0.64-1.09)	
Oral contraceptive use					
Ever/never	1.16 (0.74-1.82)	0.78 (0.63-0.96)	0.85 (0.76-0.96)	0.81 (0.69-0.95)	0.38
Duration of use, per 5 year increase	0.79 (0.62-1.02)	0.81 (0.72-0.92)	0.90 (0.84-0.96)	0.79 (0.71-0.89)	0.18
Duration of use, years					
Never	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	
≤1	1.89 (1.01-3.54)	0.96 (0.67-1.38)	1.04 (0.67-1.24)	0.99 (0.76-1.29)	
>1-≤5	1.05 (0.60-1.86)	0.89 (0.67-1.19)	0.83 (0.71-0.97)	0.89 (0.71-1.11)	0.36
>5-≤10	0.98 (0.50-1.94)	0.83 (0.60-1.16)	0.77 (0.64-0.93)	0.62 (0.46-0.84)	
>10	0.60 (0.23-1.54)	0.44 (0.27-0.73)	0.75 (0.60-0.94)	0.51 (0.35-0.75)	
Duration of breastfeeding, per 1 year[d]	1.06 (0.68-1.66)	0.93 (0.74-1.15)	0.95 (0.83-1.08)	0.89 (0.74-1.08)	0.86
Age at menarche					
Per 1 year increase	1.02 (0.91-1.14)	1.00 (0.94-1.06)	1.00 (0.97-1.03)	0.95 (0.90-0.98)	0.24
Age in years					
≤11	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	
12	1.20 (0.65-2.19)	0.93 (0.70-1.26)	1.06 (0.90-1.24)	0.86 (0.69-1.06)	
13	1.30 (0.78-2.18)	0.97 (0.75-1.26)	1.12 (0.97-1.28)	0.78 (0.64-0.95)	0.12
14	1.17 (0.58-2.32)	0.76 (0.53-1.09)	1.14 (0.96-1.37)	0.78 (0.60-1.01)	
≥15	1.01 (0.47-2.14)	1.09 (0.78-1.52)	0.87 (0.71-1.07)	0.77 (0.59-1.01)	

Age at menopause					
Per 5 year increase	1.50 (1.19-1.87)	1.03 (0.92-1.16)	1.04 (0.97-1.11)	1.02 (0.92-1.13)	0.10
Age in years					
≤45	0.23 (0.08-0.64)	0.95 (0.67-1.35)	0.91 (0.76-1.09)	0.95 (0.72-1.26)	
>45-≤50	0.46 (0.25-0.83)	1.14 (0.86-1.50)	0.96 (0.83-1.11)	1.04 (0.83-1.29)	**0.05**
>50-≤55	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	
>55	0.29 (0.07-1.18)	1.19 (0.74-1.92)	0.99 (0.77-1.28)	1.22 (0.85-1.77)	
HT use[e]					
Ever/never	1.81 (1.13-2.91)	1.68 (1.33-2.11)	1.49 (1.33-1.68)	1.29 (1.08-1.54)	0.25
Duration of use, per 5 year increase	1.35 (1.19-1.53)	1.27 (1.18-1.37)	1.21 (1.16-1.27)	1.22 (1.14-1.31)	0.54
Duration of use, years					
Never	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	
≤5	1.32 (0.69-2.54)	1.39 (1.02-1.89)	1.25 (1.07-1.46)	1.15 (1.07-1.45)	0.45
>5	3.01 (1.75-5.17)	2.19 (1.65-2.91)	1.79 (1.54-2.06)	1.66 (1.33-2.07)	
Tubal ligation, ever/never	1.26 (0.67-2.39)	1.07 (0.72-1.60)	0.93 (0.77-1.12)	0.62 (0.44-0.89)	0.11
Hysterectomy, ever/never[f]	0.88 (0.53-1.46)	1.15 (0.91-1.46)	1.04 (0.92-1.19)	1.07 (0.89-1.28)	0.79
Endometriosis, yes/no	3.77 (1.24-11.5)	1.54 (0.72-3.30)	1.11 (0.70-1.74)	0.57 (0.18-1.80)	0.12
First degree family history of breast cancer, yes/no	1.24 (0.70-2.21)	1.24 (0.94-1.67)	1.13 (0.97-1.32)	1.13 (0.97-1.32)	0.74
First degree family history of ovarian cancer, yes/no	0.90 (0.22-3.71)	1.35 (0.76-2.41)	1.61 (1.23-2.10)	1.58 (1.04-2.40)	0.80
Body mass index					
Per 5 kg/m^2	0.88 (0.71-1.11)	0.97 (0.88-1.06)	0.92 (0.87-0.97)	1.04 (0.96-1.13)	0.06
In kg/m^2					
<20	1.36 (0.69-2.69)	0.84 (0.55-1.28)	1.16 (0.96-1.41)	1.15 (0.85-1.55)	
20-<25	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	
25-<30	0.95 (0.60-1.52)	1.04 (0.83-1.29)	0.83 (0.73-1.29)	0.87 (0.73-1.04)	0.52
30-<35	0.80 (0.40-1. 95)	0.93 (0.68-1.27)	0.85 (0.71-1.00)	1.00 (0.79-1.28)	
≥35	0.98 (0.41-2.32)	0.85 (0.54-1.35)	0.89 (0.71-1.11)	1.22 (0.90-1.67)	
Height					
Per 0.5m	1.07 (0.94-1.21)	1.04 (0.97-1.12)	1.07 (1.03-1.11)	1.02 (0.96-1.08)	0.53

In meters					
<1.60	0.93 (0.53-12.60)	0.95 (0.73-1.22)	0.80 (0.70-0.93)	1.00 (0.82-1.22)	
1.60-<1.65	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	0.55
1.65-<1.70	1.38 (0.84-2.29)	0.99 (0.77-1.28)	1.03 (0.90-1.18)	1.15 (0.94-1.40)	
≥1.70	1.14 (0.64-2.04)	1.06 (0.80-1.41)	1.04 (0.90-1.21)	0.95 (0.75-1.21)	
Smoking					
Ever/never	1.14 (0.87-1.49)	0.96 (0.83-1.10)	0.95 (0.89-1.17)	1.05 (0.94-1.17)	0.30
Continuous pack-years, per 20 pack-years	0.90 (0.61-1.33)	1.00 (0.87-1.16)	0.99 (0.93-1.06)	1.08 (0.98-1.20)	0.50
Categorical pack-years					
Never	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	1.00 (ref.)	
≤20	1.21 (0.67-2.22)	1.01 (0.75-1.39)	1.06 (0.92-1.23)	1.05 (0.82-1.34)	0.95
>20	0.82 (0.37-1.78)	1.03 (0.74-1.43)	1.06 (0.90-1.24)	1.15 (0.90-1.48)	

[a]Stratified on birth year and cohort, and adjusted for age at study entry, parity, and duration of oral contraceptive use (except when parity or oral contraceptive use was the primary exposure of interest and then we adjusted only for the other risk factor) using pooled analyses of all cohorts combined. Excluding 5 cohorts with no information on grade for any ovarian cancer cases.

[b]Number of cases ranges from 29 (breastfeeding)-125 (OC use) for well-differentiated, 114 (Endometriosis)-505 (OC use) for moderately-differentiated, 343 (breastfeeding)-1669 (OC use) for poorly-differentiated, and 141 (endometriosis)-790 (OC use) for unknown grade.

[c]Assessed using a likelihood ratio test comparing a Cox proportional hazards competing risks model allowing the association to vary by grade to a model forcing the association to be the same across grades.

[d]Parous women only.

[e]Postmenopausal women only.

[f]Additionally adjusted for duration of hormone therapy use.

Appendix 2: Revised, submitted manuscript outlining the methods for risk prediction modeling in the Ovarian Cancer Association Consortium that are being applied to the OC3

Risk Prediction for Epithelial Ovarian Cancer in Eleven United States-Based Case-Control Studies: Incorporation of Epidemiologic Risk Factors and Seventeen Confirmed Genetic Loci

Running Title
Risk Prediction for Ovarian Cancer

Authors
Merlise A. Clyde[1], Rachel Palmieri Weber[2], Edwin S. Iversen[1], Elizabeth M. Poole[3], Jennifer A. Doherty[4], Marc T. Goodman[5,6], Roberta B. Ness[7], Harvey A. Risch[8], Mary Anne Rossing[9,10], Kathryn L. Terry[11,12], Nicolas Wentzensen[13], Alice S. Whittemore[14], Hoda Anton-Culver[15], Elisa V. Bandera[16], Andrew Berchuck[17], Michael E. Carney[18], Daniel W. Cramer[11,12], Julie M. Cunningham[19], Kara L. Cushing-Haugen[9], Robert P. Edwards[20,21], Brooke L. Fridley[22], Ellen L. Goode[23], Galina Lurie[24], Valerie McGuire[14], Francesmary Modugno[20,21,25], Kirsten B. Moysich[26], Sara H. Olson[27], Celeste Leigh Pearce[28,29], Malcolm C. Pike[27], Joseph H. Rothstein[14], Thomas A. Sellers[30], Weiva Sieh[14], Daniel Stram[28], Pamela J. Thompson[6], Robert A. Vierkant[31], Kristine G. Wicklund[9], Anna H. Wu[28], Argyrios Ziogas[15], Shelley S. Tworoger[3,11], and Joellen M. Schildkraut[2,32] on behalf of the Ovarian Cancer Association Consortium

Affiliations
[1] Department of Statistical Science, Duke University, Durham, North Carolina, USA
[2] Department of Community and Family Medicine, Duke University School of Medicine, Durham, North Carolina, USA
[3] Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA
[4] Department of Community and Family Medicine, Section of Biostatistics & Epidemiology, Geisel School of Medicine, Dartmouth College, Hanover, New Hampshire, USA
[5] Cancer Prevention and Control, Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, California, USA
[6] Community and Population Health Research Institute, Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, California, USA
[7] The University of Texas School of Public Health, Houston, Texas, USA
[8] Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, Connecticut, USA
[9] Program in Epidemiology, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA
[10] Department of Epidemiology, University of Washington, Seattle, Washington, USA
[11] Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA
[12] Obstetrics and Gynecology Epidemiology Center, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA
[13] Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland, USA
[14] Department of Health Research and Policy - Epidemiology, Stanford University School of Medicine, Stanford, California, USA
[15] Department of Epidemiology, School of Medicine, University of California Irvine, Irvine, California, USA

[16] Cancer Prevention and Control Program, Rutgers Cancer Institute of New Jersey, New Brunswick, New Jersey, USA
[17] Department of Obstetrics and Gynecology, Duke University School of Medicine, Durham, North Carolina, USA
[18] Department of Obstetrics and Gynecology, John A. Burns School of Medicine, University of Hawaii, Honolulu, Hawaii, USA
[19] Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota, USA
[20] Division of Gynecologic Oncology, Department of Obstetrics and Gynecology and Reproductive Sciences, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, USA
[21] Ovarian Cancer Center of Excellence, Women's Cancer Program, Magee-Womens Research Institute, University of Pittsburgh Cancer Institute, University of Pittsburgh, Pittsburgh, Pennsylvania, USA
[22] University of Kansas Medical Center, Kansas City, Kansas, USA
[23] Department of Health Science Research, Division of Epidemiology, Mayo Clinic, Rochester, Minnesota, USA
[24] Cancer Epidemiology Program, University of Hawaii Cancer Center, Honolulu, Hawaii, USA
[25] Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, USA
[26] Department of Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, New York, USA
[27] Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, New York, USA
[28] Department of Preventive Medicine, Keck School of Medicine, University of Southern California Norris Comprehensive Cancer Center, Los Angeles, California, USA
[29] Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, Michigan, USA
[30] Department of Cancer Epidemiology, Moffitt Cancer Center, Tampa, Florida, USA
[31] Department of Health Science Research, Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, Minnesota, USA
[32] Cancer Control and Population Sciences, Duke Cancer Institute, Durham, North Carolina, USA

Corresponding Author
Joellen M. Schildkraut
Box 715 Med Ctr
6036 Hock Plaza
2424 Erwin Road
Durham, NC 27705
Phone: 919-681-4761
Fax: 919-681-4766
Email: joellen.schildkraut@duke.edu

Financial Support
The scientific development and funding for this project were supported by the following: the Genetic Associations and Mechanisms in Oncology (GAME-ON): a NCI Cancer Post-GWAS Initiative (U19-CA148112); the COGS project is funded through a European Commission's Seventh Framework Programme grant (agreement number 223175-HEALTH-F2-2009-223175); the Ovarian Cancer Association Consortium is supported by a grant from the Ovarian Cancer Research Fund thanks to donations by the family and friends of Kathryn Sladek Smith (PPD/RPCI.07); and Department of Defense Awarda (W81XWH-12-1-0561 and W81XWH-12-1-

0561). F.M. was supported by National Institutes of Health K07-CA080668; S.S.T. and E.M.P. were supported in part by Department of Defense Award W81XWH-12-1-0561. Funding of the constituent studies was provided by: California Cancer Research Program (00-01389V-20170, 2II0200); Cancer Prevention Institute of California; Department of Defense (DAMD17-02-1-0666, DAMD17-02-1-0669, W81XWH-10-1-02802); Fred C. and Katherine B. Andersen Foundation; Lon V Smith Foundation grant LVS-39420; Mayo Foundation; Minnesota Ovarian Cancer Alliance; National Institutes of Health (K07-CA095666, K07-CA143047, K22-CA138563, National Center for Research Resources/General Clinical Research Center grant M01-RR000056, N01-CN025403, N01-CN55424, N01-PC67001, N01-PC67010, P01-CA17054, P30-CA14089, P30-CA15083, P30-CA072720, P50-CA105009, P50-CA136393, P50-CA159981, R01-CA058860, R01-CA074850, R01-CA080742, R01-CA092044, R01-CA112523, R01-CA122443, R01-CA126841, R01-CA16056, R01-CA54419, R01-CA58598, R01-CA61132, R01-CA76016, R01-CA83918, R01-CA87538, R01-CA95023, R03-CA113148, R03-CA115195, U01-CA69417, U01-CA71966); Rutgers Cancer Institute of New Jersey; and the US Public Health Service (PSA-042205).

Conflicts of Interest
No potential conflicts of interest were disclosed by the authors.

ABSTRACT

Previously developed models for predicting absolute risk of invasive epithelial ovarian cancer have considered a limited number of risk factors and have low discriminatory power (area under the receiver operating characteristic curve, AUCs<0.60). As such, we developed and internally validated a relative risk prediction model that incorporates 17 established epidemiological risk factors and 17 genome-wide significant single nucleotide polymorphisms (SNPs) using data from 11 case-control studies in the United States (5,793 cases; 9,512 controls) from the Ovarian Cancer Association Consortium. We developed a hierarchical logistic regression model for predicting case-control status that included imputation of missing data. We randomly divided the data into an 80% training sample and used the remaining 20% for model evaluation. The AUC for the full model was 0.664. A reduced model without SNPs performed similarly (AUC=0.649). Both models performed better than a baseline model with age and study site only (AUC=0.563). The best predictive power was obtained in the full model among women under 50 years of age (AUC=0.714), however, the addition of SNPs increased the AUC the most for women over 50 (AUC = 0.638 versus 0.616). Adapting this improved model to estimate absolute risk and evaluating it in prospective datasets is warranted.

Introduction
Almost 22,000 new cases of ovarian cancer and 14,270 deaths from ovarian cancer were expected in 2014, accounting for 5% of cancer deaths among women; most (85-90%) are epithelial (1). The five-year survival for localized ovarian cancer is 92%, but most cases are diagnosed at a distant stage when the five-year survival is only 27% (2). Epithelial ovarian cancer (EOC) has no specific symptoms, and no screening or early detection measures have been adopted clinically, making disease prevention and identification of high-risk women key to reducing mortality (1).

Risk prediction models can provide objective estimates for use in clinical decision-making, identification of highest-risk individuals who can benefit from preventive measures,

development of preventive intervention studies at the population level, and creation of risk-benefit indices (3). EOC risk prediction is challenging due to its rarity and the modest effects of most known risk factors, although several well-established risk factors have been identified. Oral contraceptive (OC) use (4), parity (5), and tubal ligation (6, 7) are inversely associated with risk of EOC; family history of breast or ovarian cancer is positively associated with risk (8). Older age at menarche and menopausal hormone therapy (MHT) (particularly estrogen only therapy) have been associated with increased risk of EOC while breastfeeding and hysterectomy have been associated with decreased risk, in some, but not all, studies (6, 9-16). Although reports have been inconsistent, a recent report of 12 population-based case-control studies concluded that aspirin use was associated with reduced EOC risk (17). Further, endometriosis has been associated with risk of low-grade serous, endometrioid, and clear cell EOC (18, 19).

EOC risk prediction models generally have low discrimination (area under the curve (AUC) <0.60), which may be partly due to exclusion of women who reported premenopausal hysterectomy (with or without unilateral oophorectomy), incomplete inclusion of risk factors (e.g., tubal ligation), or prediction in specific sub-populations (e.g., at time of hysterectomy or women with symptoms) (20-25). Although some existing risk prediction models specifically address risk among BRCA1 and BRCA2 mutation carriers (26, 27), these mutations are rare in the general population; prior models for women of average risk have not considered genetic susceptibility. With 17 confirmed genetic susceptibility variants reported for EOC (28-34), our objective was to develop and internally validate a relative risk prediction model for invasive EOC among women of average risk that incorporated all established and strongly probable epidemiologic risk factorsand genetic susceptibility data from 11 case-control studies in the United States (US) that are members of the Ovarian Cancer Association Consortium (OCAC).

METHODS

Study populations and inclusion criteria

The analysis included 11 US-based case-control studies in the OCAC (Table 1) (14, 35-45). All studies were population-based, with the exception of the MAY study, which was clinic-based; MAY controls were women attending the Mayo Clinic's Departments of Family Medicine and General Internal Medicine for general medical exams. All studies had ethics board approval and obtained written informed consent. Data were included for women who were 30 years of age or older at diagnosis (cases) or interview/reference date (controls), had no prior history of cancer (except non-melanoma skin cancer), and self-identified as white, non-Hispanic; most women were confirmed to be of European ancestry by genetic analysis. Controls had to have at least one intact ovary and cases were limited to invasive EOC. Most cases (81%) were recruited within one year of diagnosis. After exclusions, the analysis included data from 5,793 invasive EOC case patients and 9,512 controls. We randomly sampled 80% of the participants (n=12,244) for estimation and model building; the remaining 20% (n=3,061) were retained for independent validation.

Risk factor data

Data from each study on known and suspected risk factors, and demographic and clinical variables, were submitted to the OCAC data coordination center at Duke, where common coding schemes were applied; data were originally collected via questionnaire. The following risk factors were available in the majority of studies: age at menarche (continuous years); OC use (ever/never); duration of OC use (continuous months); aspirin use (low dose, high dose, or irregular/no use); number of full term pregnancies (continuous), number of non-full term pregnancies (continuous variable; derived by subtracting parity from number of pregnancies); breastfeeding status (ever/never); duration of breastfeeding (continuous months); age at end of last pregnancy (continuous years); tubal ligation (yes/no); hysterectomy more than 1 year prior to diagnosis (cases) or interview/reference age (controls) (yes/no); endometriosis

(yes/no); body mass index (BMI) within five years of diagnosis/interview; menopause status at diagnosis (cases) or interview/reference age (controls) (pre-/post-menopausal); MHT use (ever/never); type of MHT (unopposed estrogen replacement therapy only/all other MHT use); history of breast cancer in a first degree relative (yes/no); and history of ovarian cancer in a first degree relative (yes/no). We considered additional potential risk factors (e.g., non-steroidal anti-inflammatory drug use (NSAIDs), age at tubal ligation) that were ultimately not included because they were not significant predictors of EOC in preliminary models and were missing from a large percentage of participants. Due to frequency matching, age was included in all models to avoid bias (46).

Genetic susceptibility data

The OCAC evaluated 23,239 SNPs in 43 individual studies that were grouped into 34 case-control strata; two previous genome-wide association studies (GWAS) informed the OCAC-specific SNP selection for the Collaborative Oncological Gene-environment Study (COGS) (34). Analysis of the GWAS and COGS genotype data identified and confirmed 17 susceptibility loci (Supplemental Table 1) (28-34) that are included in our risk prediction model. Some, but not all, participants from the studies in our analysis contributed to the GWAS (MAY, NCO, NEC) and COGS (all studies except CON) genotyping efforts, requiring imputation of missing SNPS for the remaining women.

Statistical Analysis

We used generalized additive models (GAMs) (R package *mgcv*) (47-49) with random effects for study site, fixed effects for categorical variables and SNPs, and smooth non-parametric functions for continuous variables as part of exploratory model fitting using subjects with complete data. Some evidence supports that risk factor associations may vary by menopausal status (50). However, because age at menopause was missing from 59% of the post-menopausal women and is difficult to determine for some women due to premenopausal hysterectomy and hormone use, we fit separate models for women under 50 years of age and women 50 years and older. The GAMs suggested that nonlinear functions of the continuous variables could be approximated with linear functions of the variables ($p > 0.05$) with the exception of OC duration. The square root of OC duration did not produce a significant increase in the deviance over using the spline terms ($p = 0.2265$), while a linear term for OC duration was rejected ($p=0.0114$). We retained linear terms with the original continuous variables except for OC duration, which used the square root transformation.

All risk factors except age had some missing data; 80% of the participants were missing information on at least one risk factor (Table 2). Rather than limit analysis to participants with complete data or drop risk factors from the model, we developed a Bayesian model (51) that provided a coherent sequence of conditional models for case-control status, the risk factors, and indicators of whether they are missing (in the case of data not missing at random) (52); missing risk factors and indicators were modeled as functions of other risk covariates as well as education level, smoking status, and alcohol use (Table 3). The joint model specification for the risk factors and ovarian cancer status allowed all observed data to be incorporated and simultaneous inference for model parameters and missing data via Markov Chain Monte Carlo (MCMC) using JAGS (53). The increased sample size obtained by using participants with partial information can increase power, while the multiple imputations through MCMC provide valid confidence intervals for statistical inference by addressing uncertainty in the missing values and reducing bias induced by complete case analyses when data are not missing at random (54).

The first stage Bernoulli models expressed the log odds of the probability of EOC (π_i)

(i) $$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \alpha^{g}_{\text{site}_i} + \sum_{j=1}^{6} Z_{i,j}\beta^{g}_{j,c_i} + \sum_{j} X_{i,j}\beta^{g}_{j}$$

for the two groups (denoted by *g*) via a generalized linear mixed model with random effects for the 11 studies to account for differential baseline odds due to study design:

(ii) $$\alpha^{g}_{\text{site}_i} \overset{ind}{\sim} N(\mu_{\text{site}}, \sigma^2_{site})$$

random effects to account for birth cohort (*c*):

(iii) $$\beta^{g}_{j,c_i} \overset{ind}{\sim} N(\beta^{g}_{j}, \sigma^2_{j})$$

for the six hormonally-related covariates *Z* (i.e., indicator of OC use, square root of OC duration, indicator of MHT use, indicator of type of MHT usage, interaction of the indicator of hysterectomy with MHT use and type of MHT use) to allow potential birth year differences due to formulation changes, and finally fixed effects for the remaining risk factors in *X* in each group (17 epidemiological risk factors and the 17 SNPs). All of the group specific means, β^{g}_{j}, for random effects and fixed effect coefficients for the other exposures were given independent normal prior distributions, with a mean β_j and a prior standard deviation of one, reflecting the expectation that population log odds ratios (log ORs) should be well within plus or minus 2 based on prior estimates and standard deviations from the literature. For the 17 SNPs, we used informative prior distributions based on log ORs from the GWAS and COGS samples independent from the 11 studies included in model development (Supplemental Table 2). The hierarchical formulation allows coefficients to "shrink" to common coefficients across sites, cohorts and age groups if significant variation is not present, but provides flexibility to account for differences among groups while avoiding issues of multiple testing. Distributions for the missing data models are given in Table 3. For example, missing SNPs were modeled using a multinomial model with the probabilities for the number of rare alleles given an informative Dirichlet prior distribution centered at genotype probabilities under Hardy-Weinberg and a mass parameter in the Dirichlet equivalent to 1000 observations; genotype probabilities were calculated using the Minor Allele Frequencies (MAF) taken from GWAS and COGS samples from OCAC not used in this analysis (Supplemental Table 2). Combined with genotype data, other risk variables, and case-control status, missing SNPs were generated using their respective predictive distributions given the observed data and values of parameters at each iteration in the Markov chain.

Models with and without the SNPs were fit to the training data (random sample of 80%) and used to predict case-control status on the validation data (remaining 20%). Inference was based on 70,000 iterations of the MCMC algorithm. The first 20,000 iterations were used to assess convergence of the MCMC and the last 50,000 were used for inference with the training data and predictions in the validation set. Point estimates of log ORs were estimated by the median of the samples from the posterior distribution of each of the parameters; (Bayesian) 95% confidence intervals (CI) were obtained by taking the 2.5th percentile and 97.5th percentile of the estimated posterior distribution for each parameter (55). Predictions for each participant in the training data were based on the mean of the posterior predictive distribution which was estimated using the Monte Carlo average over posterior draws of missing predictors and parameters in equation (i). For comparison, we also fit a model adjusting for study site and age only (baseline model), and study site, age, and SNPs, omitting the epidemiological risk factors.

Model validation

We compared the models with and without SNPs, and with and without the epidemiological variables, on the basis of their overall discriminatory accuracy and calibration in the independent validation data. We evaluated the discriminatory accuracy of the risk prediction models using the AUC from the receiver operating characteristics (ROC) curve. Predictive performance on the validation set was also assessed using calibration plots that compared the

predicted risk (score) from the model to the observed proportions across groups defined by study sites, birth cohorts, age, and number of pregnancies.

RESULTS

The training set had 4,662 cases and 7,586 controls; the evaluation set had 1,131 cases and 1,926 controls (Table 2). Women averaged 57 years of age. As expected, in both the training and evaluation sets, case patients were less likely to use OCs, have been pregnant, and have a tubal ligation than controls and more likely to have a family history of breast or ovarian cancer and use MHT. The distribution of SNPs was similar to those observed in the larger GWAS and COGs datasets.

Table 4 provides estimates of the log ORs (medians) and 95% Bayesian CIs for the group-specific coefficients from the hierarchical logistic regression model with the 17 SNPs; estimates from the model without the 17 SNPs were similar (Supplemental Table 3). Most of the epidemiological risk factors included in the model were statistically significant predictors among women under 50, however, in general, the directions of associations were comparable across groups. Notably, some associations were weaker among older women compared to the younger women, including duration of OC use, number of pregnancies and breastfeeding, family history of breast or ovarian cancers, endometriosis, tubal ligation, MHT use and type, and hysterectomy, while low-dose aspirin use showed a significant protective effect in women age 50 and older. Furthermore more of the SNPs were statistically significant for women age 50 and older, which represent the majority of women in this study. Endometriosis, duration of OC use, tubal ligation, family history of breast and ovarian cancer, number of non-full term pregnancies, rs2072590, rs10088218 in 8q24, rs9303542, rs7651446 in 5p15, rs3814113, rs56318008, and rs183211 contributed significantly to all of the group-specific models.

The AUC for models for all women, women under 50, and women 50 and over, for the models without and with SNPs are shown in Figures 1A and 1B, respectively; the inclusion of the SNPs provided a small improvement (0.015 change in the AUC) in predictions for the validation data in terms of AUC for all women, with the biggest improvement for women 50 and over (0.026 increase). Among all women, the AUC was 0.664 with SNPs and 0.649 without SNPs (but including epidemiological factors), which is a marked improvement over the AUC for the models with age and study site alone (AUC=0.563) and for age, study site, and the 17 SNPs (AUC=0.600) (Table 5). The posterior probability that the AUC for the full model with SNPs and epidemiological factors is better than the AUC for the model with age, study site, and SNPs alone was 99.8%, while there was a 70% chance that the addition of SNPs improved AUC over the model with age, study site, and epidemiological factors. The best predictive power was obtained for women under 50: AUC=0.714 and 0.713 in the models with and without the SNPs, respectively. Lower AUCs were observed in women 50 and over (0.638 with SNPs and 0.612 without SNPs). Finally for comparison, we generated a target ROC curve with an AUC of 0.75 for a widely accepted clinically actionable discrimination by sequentially adding hypothetical SNPs generated with a minor allele frequency of 0.20 and a log odds ratio of 0.15 (within the range of currently validated SNPS for EOC) until the AUC exceeded 0.75. Under this setting, on average 58 additional SNPS would be needed (95% CI: 39, 79) to increase the AUC from 0.66 to 0.75.

Figure 2 and Supplemental Figure 1 suggest that the model is well-calibrated across risk score deciles, studies, birth cohorts, age, and number of pregnancies.

DISCUSSION

Our validated relative risk prediction model for EOC includes an extensive list of established non-genetic risk factors for ovarian cancer and 17 novel genetic variants. We divided the data set of 5,793 cases and 9,512 controls of non-Hispanic, European ancestry, in an 80:20 ratio for use in independent modeling and evaluation analyses. Overall, the model's

predictive capacity was modest and epidemiologic factors contributed to the increase in the AUC substantially more than the SNPs. The methodology for imputation developed here may be adapted for prospective validation.

Previous ovarian cancer risk prediction analyses have included fewer than 1,000 cases in any given phase of model development or validation (23, 24). Our much larger sample size provided ample power for stratification by age (<50 versus ≥50) and permitted us to include a much larger number of accepted epidemiologic risk factors, as well as 17 genetic loci. This, coupled with the imputation of missing data, provided the power necessary to detect and estimate higher order interaction effects. The model includes an interaction between MHT use and hysterectomy status dependent on age.

In contrast to previous models, we developed a joint model for disease status, risk factors, and missingness. A strength of our approach was the use of MCMC methods that allow for simultaneous inference for missing data and model parameters. This allowed us to include all participants in the analysis while correctly accounting for the observed sample sizes in interval and error estimates of odds ratios. This is critical when variables, such as hysterectomy status, are not missing at random and would therefore lead to biased inferences using most standard methods, including complete-case analysis (54). The hierarchical framework also permits parsimonious adjustment for birth cohort effects in hormonal exposures, such as OC and MHT use, where formulations have changed over time.

To date, absolute risk prediction models for ovarian cancer have achieved moderate discriminatory accuracy in the general population. A recent model, which included first degree family history of breast or ovarian cancer, duration of MHT use, parity, and duration of OC use, and was developed and externally validated among women over age 50, had an AUC of 0.59 (23). The best model from the Nurses' Health Studies included duration of ovulation (age (for premenopausal women) or age at menopause minus age at menarche minus one year per pregnancy and years of OC use), duration of menopause, and tubal ligation; the overall AUC for the model predicting ovarian cancer was approximately 0.60 (24). Our full model obtained higher overall predictive accuracy (AUC=0.664), albeit estimated in a case-control setting, in part because more established risk factors were included and we allowed for associations to vary by strata in the population (age), as well as birth cohorts.

The predictive ability of the model was substantially higher for younger (AUC=0.714) than older women (AUC=0.638), despite the increase in incidence of ovarian cancer with age. This is consistent with the Rosner risk prediction model (24), in which the AUCs generally were higher for women under 50. One reason for the improved prediction in younger women is that many of the risk factors occur during pre-menopause and appear to have stronger associations in younger women, perhaps in part because the exposure to the risk factors is more proximal (50). Our results are consistent with studies of individual risk factors suggesting, for example, that the protective effects of hysterectomy, OC use and tubal ligation attenuate with increasing time since last use (or surgery) (4, 6, 50).

Recent efforts to improve risk estimation have focused on common genetic variation. However, the addition of common SNPs to risk prediction models has not yet resulted in dramatically improved discriminatory accuracy, in real or simulated data scenarios (56-58). Our findings are consistent with this; addition of the 17 confirmed SNPs improved the AUC of the model incorporating epidemiologic risk factors by a small amount (AUC=0.664 with SNPs versus AUC=0.649 without). Our model addresses women of average baseline risk and mutation status of highly penetrant susceptibility genes such as BRCA1 and BRCA2 was not included since these data were not available. Although the model accounts for family history of breast and ovarian cancer, the inclusion of the mutation status and other high penetrant rare variants may improve prediction in future efforts. However, even strongly associated risk factors (genetic or non-genetic) may only modestly improve upon a risk model's discriminatory accuracy (59) and a very large number of susceptibility SNPs (i.e., several hundred) are

required to make a substantial impact because of their small relative risks (60); our simulation results suggest that an additional 39 to 79 SNPs may be needed to increase the AUC to a clinically actionable discriminatory value of 0.75. This is similar to observations for breast cancer, where a 3-4 unit increase can be achieved with addition of 60-70 SNPs (61-66).

The model may be improved by extension to predict histologic subtypes of EOC, as risk factor associations may vary by histology (19). Further gains in predictive accuracy may accompany discovery and inclusion of additional novel risk factors. In breast cancer, the addition of sex hormones and mammographic density added substantially to risk prediction models (67, 68). Finally, these results may not be generalizable to other racial or ethnic groups or to other countries.

Our model was developed and internally validated among participants from case-control studies. Although this study design may be subject to misclassification and selection bias, the studies were predominantly population-based and our associations are similar in direction and magnitude to those observed in cohort studies. To be clinically meaningful, the relative risk estimates must be combined with a model of age-specific baseline population risk to provide estimates of absolute risk. Hierarchical models provide a natural framework for integrating relative risk estimates from this study -- and propagating their uncertainty -- into future models for absolute risk within prospective studies.

REFERENCES

1. American Cancer Society. Cancer Facts & Figures 2014. Atlanta: American Cancer Society; 2014.
2. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. CA: a cancer journal for clinicians 2014;64(1):9-29.
3. Freedman AN, Seminara D, Gail MH, Hartge P, Colditz GA, Ballard-Barbash R, et al. Cancer risk prediction models: a workshop on development, evaluation, and application. Journal of the National Cancer Institute 2005;97(10):715-23.
4. Beral V, Doll R, Hermon C, Peto R, Reeves G. Ovarian cancer and oral contraceptives: collaborative reanalysis of data from 45 epidemiological studies including 23,257 women with ovarian cancer and 87,303 controls. Lancet 2008;371(9609):303-14.
5. Adami HO, Hsieh CC, Lambe M, Trichopoulos D, Leon D, Persson I, et al. Parity, age at first childbirth, and risk of ovarian cancer. Lancet 1994;344(8932):1250-4.
6. Rice MS, Murphy MA, Tworoger SS. Tubal ligation, hysterectomy and ovarian cancer: A meta-analysis. Journal of ovarian research 2012;5(1):13.
7. Sieh W, Salvador S, McGuire V, Weber RP, Terry KL, Rossing MA, et al. Tubal ligation and risk of ovarian cancer subtypes: a pooled analysis of case-control studies. International journal of epidemiology 2013;42(2):579-89.
8. Schildkraut JM, Risch N, Thompson WD. Evaluating genetic association among ovarian, breast, and endometrial cancer: evidence for a breast/ovarian cancer relationship. American journal of human genetics 1989;45(4):521-9.
9. Beral V, Bull D, Green J, Reeves G. Ovarian cancer and hormone replacement therapy in the Million Women Study. Lancet 2007;369(9574):1703-10.
10. Coughlin SS, Giustozzi A, Smith SJ, Lee NC. A meta-analysis of estrogen replacement therapy and risk of epithelial ovarian cancer. Journal of clinical epidemiology 2000;53(4):367-75.
11. Gong TT, Wu QJ, Vogtmann E, Lin B, Wang YL. Age at menarche and risk of ovarian cancer: a meta-analysis of epidemiological studies. International journal of cancer Journal international du cancer 2013;132(12):2894-900.

12. Luan NN, Wu QJ, Gong TT, Vogtmann E, Wang YL, Lin B. Breastfeeding and ovarian cancer risk: a meta-analysis of epidemiologic studies. The American journal of clinical nutrition 2013;98(4):1020-31.
13. Pearce CL, Chung K, Pike MC, Wu AH. Increased ovarian cancer risk associated with menopausal estrogen therapy is reduced by adding a progestin. Cancer 2009;115(3):531-9.
14. Rossing MA, Cushing-Haugen KL, Wicklund KG, Doherty JA, Weiss NS. Menopausal hormone therapy and risk of epithelial ovarian cancer. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 2007;16(12):2548-56.
15. Trabert B, Wentzensen N, Yang HP, Sherman ME, Hollenbeck A, Danforth KN, et al. Ovarian cancer and menopausal hormone therapy in the NIH-AARP diet and health study. British journal of cancer 2012;107(7):1181-7.
16. Zhou B, Sun Q, Cong R, Gu H, Tang N, Yang L, et al. Hormone replacement therapy and ovarian cancer risk: a meta-analysis. Gynecologic oncology 2008;108(3):641-51.
17. Trabert B, Ness RB, Lo-Ciganic WH, Murphy MA, Goode EL, Poole EM, et al. Aspirin, nonaspirin nonsteroidal anti-inflammatory drug, and acetaminophen use and risk of invasive epithelial ovarian cancer: a pooled analysis in the Ovarian Cancer Association Consortium. Journal of the National Cancer Institute 2014;106(2):djt431.
18. Heidemann LN, Hartwell D, Heidemann CH, Jochumsen KM. The relation between endometriosis and ovarian cancer - a review. Acta obstetricia et gynecologica Scandinavica 2014;93(1):20-31.
19. Pearce CL, Templeman C, Rossing MA, Lee A, Near AM, Webb PM, et al. Association between endometriosis and risk of histological subtypes of ovarian cancer: a pooled analysis of case-control studies. The Lancet Oncology 2012;13(4):385-94.
20. Andersen MR, Goff BA, Lowe KA, Scholler N, Bergan L, Drescher CW, et al. Use of a Symptom Index, CA125, and HE4 to predict ovarian cancer. Gynecologic oncology 2010;116(3):378-83.
21. Collins GS, Altman DG. Identifying women with undetected ovarian cancer: independent and external validation of QCancer((R)) (Ovarian) prediction model. European journal of cancer care 2013;22(4):423-9.
22. Hartge P, Whittemore AS, Itnyre J, McGowan L, Cramer D. Rates and risks of ovarian cancer in subgroups of white women in the United States. The Collaborative Ovarian Cancer Group. Obstetrics and gynecology 1994;84(5):760-4.
23. Pfeiffer RM, Park Y, Kreimer AR, Lacey JV, Jr., Pee D, Greenlee RT, et al. Risk prediction for breast, endometrial, and ovarian cancer in white women aged 50 y or older: derivation and validation from population-based cohort studies. PLoS medicine 2013;10(7):e1001492.
24. Rosner BA, Colditz GA, Webb PM, Hankinson SE. Mathematical models of ovarian cancer incidence. Epidemiology (Cambridge, Mass) 2005;16(4):508-15.
25. Vitonis AF, Titus-Ernstoff L, Cramer DW. Assessing ovarian cancer risk when considering elective oophorectomy at the time of hysterectomy. Obstetrics and gynecology 2011;117(5):1042-50.
26. Chen S, Iversen ES, Friebel T, Finkelstein D, Weber BL, Eisen A, et al. Characterization of BRCA1 and BRCA2 mutations in a large United States sample. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2006;24(6):863-71.
27. Lee AJ, Cunningham AP, Kuchenbaecker KB, Mavaddat N, Easton DF, Antoniou AC. BOADICEA breast cancer risk prediction model: updates to cancer incidences, tumour pathology and web interface. British journal of cancer 2014;110(2):535-45.

28. Bojesen SE, Pooley KA, Johnatty SE, Beesley J, Michailidou K, Tyrer JP, et al. Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. Nature genetics 2013;45(4):371-84, 84e1-2.
29. Bolton KL, Tyrer J, Song H, Ramus SJ, Notaridou M, Jones C, et al. Common variants at 19p13 are associated with susceptibility to ovarian cancer. Nature genetics 2010;42(10):880-4.
30. Goode EL, Chenevix-Trench G, Song H, Ramus SJ, Notaridou M, Lawrenson K, et al. A genome-wide association study identifies susceptibility loci for ovarian cancer at 2q31 and 8q24. Nature genetics 2010;42(10):874-9.
31. Permuth-Wey J, Lawrenson K, Shen HC, Velkova A, Tyrer JP, Chen Z, et al. Identification and molecular characterization of a new ovarian cancer susceptibility locus at 17q21.31. Nature communications 2013;4:1627.
32. Pharoah PD, Tsai YY, Ramus SJ, Phelan CM, Goode EL, Lawrenson K, et al. GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. Nature genetics 2013;45(4):362-70, 70e1-2.
33. Song H, Ramus SJ, Tyrer J, Bolton KL, Gentry-Maharaj A, Wozniak E, et al. A genome-wide association study identifies a new ovarian cancer susceptibility locus on 9p22.2. Nature genetics 2009;41(9):996-1000.
34. Kuchenbaecker KB, Ramus SJ, Tyrer J, Lee A, Shen HC, Beesley J, et al. Identification of six new susceptibility loci for invasive epithelial ovarian cancer. Nature genetics 2015;47(2):164-71.
35. Bandera EV, King M, Chandran U, Paddock LE, Rodriguez-Rodriguez L, Olson SH. Phytoestrogen consumption from foods and supplements and epithelial ovarian cancer risk: a population-based case control study. BMC women's health 2011;11:40.
36. Kelemen LE, Sellers TA, Schildkraut JM, Cunningham JM, Vierkant RA, Pankratz VS, et al. Genetic variation in the one-carbon transfer pathway and ovarian cancer risk. Cancer research 2008;68(7):2498-506.
37. Lo-Ciganic WH, Zgibor JC, Bunker CH, Moysich KB, Edwards RP, Ness RB. Aspirin, nonaspirin nonsteroidal anti-inflammatory drugs, or acetaminophen and risk of ovarian cancer. Epidemiology (Cambridge, Mass) 2012;23(2):311-9.
38. Lurie G, Wilkens LR, Thompson PJ, McDuffie KE, Carney ME, Terada KY, et al. Combined oral contraceptive use and epithelial ovarian cancer risk: time-related effects. Epidemiology (Cambridge, Mass) 2008;19(2):237-43.
39. McGuire V, Felberg A, Mills M, Ostrow KL, DiCioccio R, John EM, et al. Relation of contraceptive and reproductive history to ovarian cancer risk in carriers and noncarriers of BRCA1 gene mutations. American journal of epidemiology 2004;160(7):613-8.
40. Pike MC, Pearce CL, Peters R, Cozen W, Wan P, Wu AH. Hormonal factors and the risk of invasive ovarian cancer: a population-based case-control study. Fertility and sterility 2004;82(1):186-95.
41. Risch HA, Bale AE, Beck PA, Zheng W. PGR +331 A/G and increased risk of epithelial ovarian cancer. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 2006;15(9):1738-41.
42. Schildkraut JM, Iversen ES, Wilson MA, Clyde MA, Moorman PG, Palmieri RT, et al. Association between DNA damage response and repair genes and risk of invasive serous ovarian cancer. PloS one 2010;5(4):e10061.
43. Terry KL, De Vivo I, Titus-Ernstoff L, Shih MC, Cramer DW. Androgen receptor cytosine, adenine, guanine repeats, and haplotypes in relation to ovarian cancer risk. Cancer research 2005;65(13):5974-81.

44. Wu AH, Pearce CL, Tseng CC, Templeman C, Pike MC. Markers of inflammation and risk of ovarian cancer in Los Angeles County. International journal of cancer Journal international du cancer 2009;124(6):1409-15.
45. Ziogas A, Gildea M, Cohen P, Bringman D, Taylor TH, Seminara D, et al. Cancer risk estimates for family members of a population-based family registry for breast and ovarian cancer. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 2000;9(1):103-11.
46. Janes H, Pepe MS. Adjusting for covariates in studies of diagnostic, screening, or prognostic markers: an old concept in a new setting. American journal of epidemiology 2008;168(1):89-97.
47. R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.
48. Wood SN. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society (B) 2011;73(1):3-36.
49. Wood SN. Generalized Additive Models: An Introduction with R. Chapman and Hall/CRC; 2006.
50. Moorman PG, Calingaert B, Palmieri RT, Iversen ES, Bentley RC, Halabi S, et al. Hormonal risk factors for ovarian cancer in premenopausal and postmenopausal women. American journal of epidemiology 2008;167(9):1059-69.
51. Gelman A. Data analysis using regression and multilevel/hierarchical models. Hill J, editor. Cambridge ;New York: Cambridge University Press; 2007.
52. Little RJA. Statistical analysis with missing data. Rubin DB, editor. Hoboken, N.J.: Wiley; 2002.
53. Plummer M. JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling, Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003). March 20-22; Vienna, Austria.
54. Greenland S, Finkle WD. A critical look at methods for handling missing covariates in epidemiologic regression analyses. American journal of epidemiology 1995;142(12):1255-64.
55. Hoff PD. A first course in Bayesian statistical methods [electronic resource]. New York ;London: Springer; 2009.
56. Husing A, Canzian F, Beckmann L, Garcia-Closas M, Diver WR, Thun MJ, et al. Prediction of breast cancer risk by genetic risk factors, overall and by hormone receptor status. Journal of medical genetics 2012;49(9):601-8.
57. Park JH, Gail MH, Greene MH, Chatterjee N. Potential usefulness of single nucleotide polymorphisms to identify persons at high cancer risk: an evaluation of seven common cancers. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2012;30(17):2157-62.
58. Wacholder S, Hartge P, Prentice R, Garcia-Closas M, Feigelson HS, Diver WR, et al. Performance of common genetic variants in breast-cancer risk models. The New England journal of medicine 2010;362(11):986-93.
59. Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. American journal of epidemiology 2004;159(9):882-90.
60. Gail MH. Discriminatory accuracy from single-nucleotide polymorphisms in models to predict breast cancer risk. Journal of the National Cancer Institute 2008;100(14):1037-41.

61. Darabi H, Czene K, Zhao W, Liu J, Hall P, Humphreys K. Breast cancer risk prediction and individualised screening based on common genetic variation and breast density measurement. Breast cancer research : BCR 2012;14(1):R25.
62. Husing A, Canzian F, Beckmann L, Garcia-Closas M, Diver WR, Thun MJ. Prediction of breast cancer risk by genetic risk factors, overall and by hormone receptor status. Journal of medical genetics 2012;49(9):601-8.
63. Mealiffe ME, Stokowski RP, Rhees BK, Prentice RL, Pettinger M, Hinds DA. Assessment of clinical validity of a breast cancer risk model combining genetic and clinical information. Journal of the National Cancer Institute 2010;102(21):1618-27.
64. Pharoah PD, Antoniou AC, Easton DF, Ponder BAJ. Polygenes, risk prediction, and targeted prevention of breast cancer. The New England journal of medicine 2008;358(26):2796-803.
65. Wacholder S, Hartge P, Prentice R, Garcia-Closas M, Feigelson HS, Diver WR. Performance of common genetic variants in breast-cancer risk models. The New England journal of medicine 2010;362(11):986-93.
66. Gail MH. Value of adding single-nucleotide polymorphism genotypes to a breast cancer risk model. Journal of the National Cancer Institute 2009;101(13):959-63.
67. Tice JA, Cummings SR, Smith-Bindman R, Ichikawa L, Barlow WE, Kerlikowske K. Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model. Annals of internal medicine 2008;148(5):337-47.
68. Tworoger SS, Zhang X, Eliassen AH, Qian J, Colditz GA, Willett WC, et al. Inclusion of Endogenous Hormone Levels in Risk Prediction Models of Postmenopausal Breast Cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2014.

Figure 1. Receiver Operating Characteristic Curve for Models A) Without and B) With SNPs

The receiver operating characteristic (ROC) curve plots the true positive fraction (i.e., sensitivity) versus the false positive fraction (i.e., 1-specificity) at various threshold settings. The ROC curve in (a) represents the relative risk prediction model containing age, study site, and 17 risk factors; the ROC curve in (b) represents the full relative risk prediction model containing the variables in (a) plus 17 confirmed genetic susceptibility variants. For each model, 3 ROC curves are presented for women grouped by: all ages (dark blue), women under 50 years of age (light blue), and women 50 years of age and older (green). The area under the curve (AUC), a measure of discriminatory power equivalent to the '*c* statistic' in binary models, is presented for each ROC curve. A fourth hypothetical target ROC curve (magenta) is depicted based on adding additional hypothetical SNPs with a MAF of 0.20 and log odds ratio of 0.15 (similar to the current data)

until the AUC is 0.75 or more; on average 58 additional SNPS would be needed (95% CI: 39, 79).

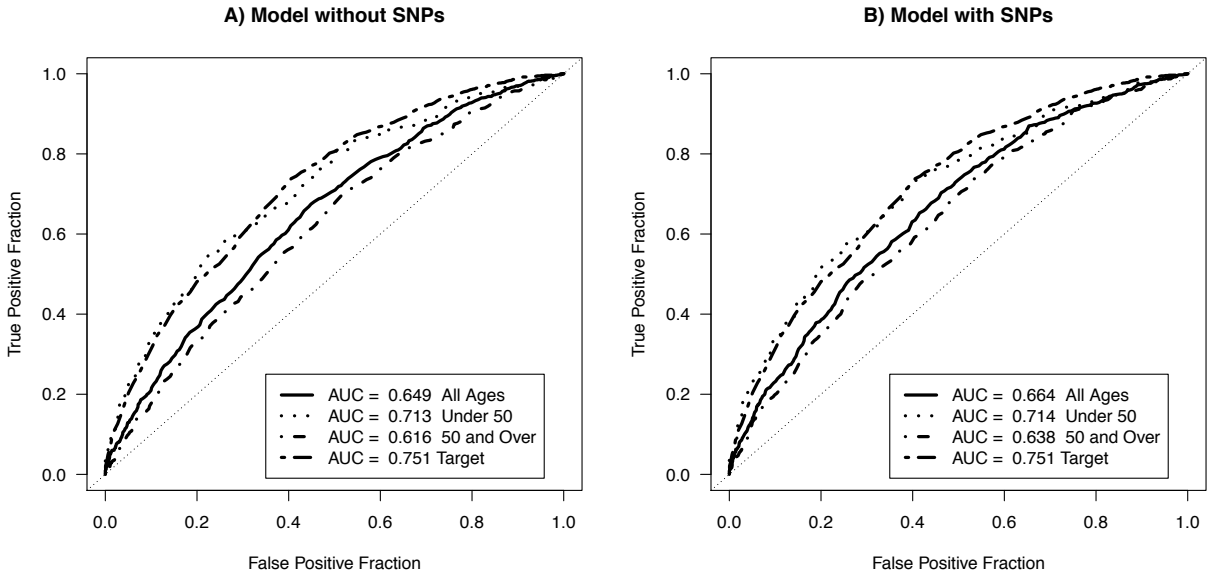


Figure 2. Calibration Plots for Risk Scores

The calibration plot represents the agreement between the average predicted probability of epithelial ovarian cancer (i.e., risk score) and observed outcomes (i.e., relative frequency of cases) in the full risk prediction model containing age, study site, 17 risk factors, and 17 confirmed genetic susceptibility variants for women included in the analysis. Women were divided into ten bins determined by increasing risk (0.10 long). The vertical and horizontal bars reflect uncertainty in the average predicted risk and mean under a Bernoulli model, respectively.

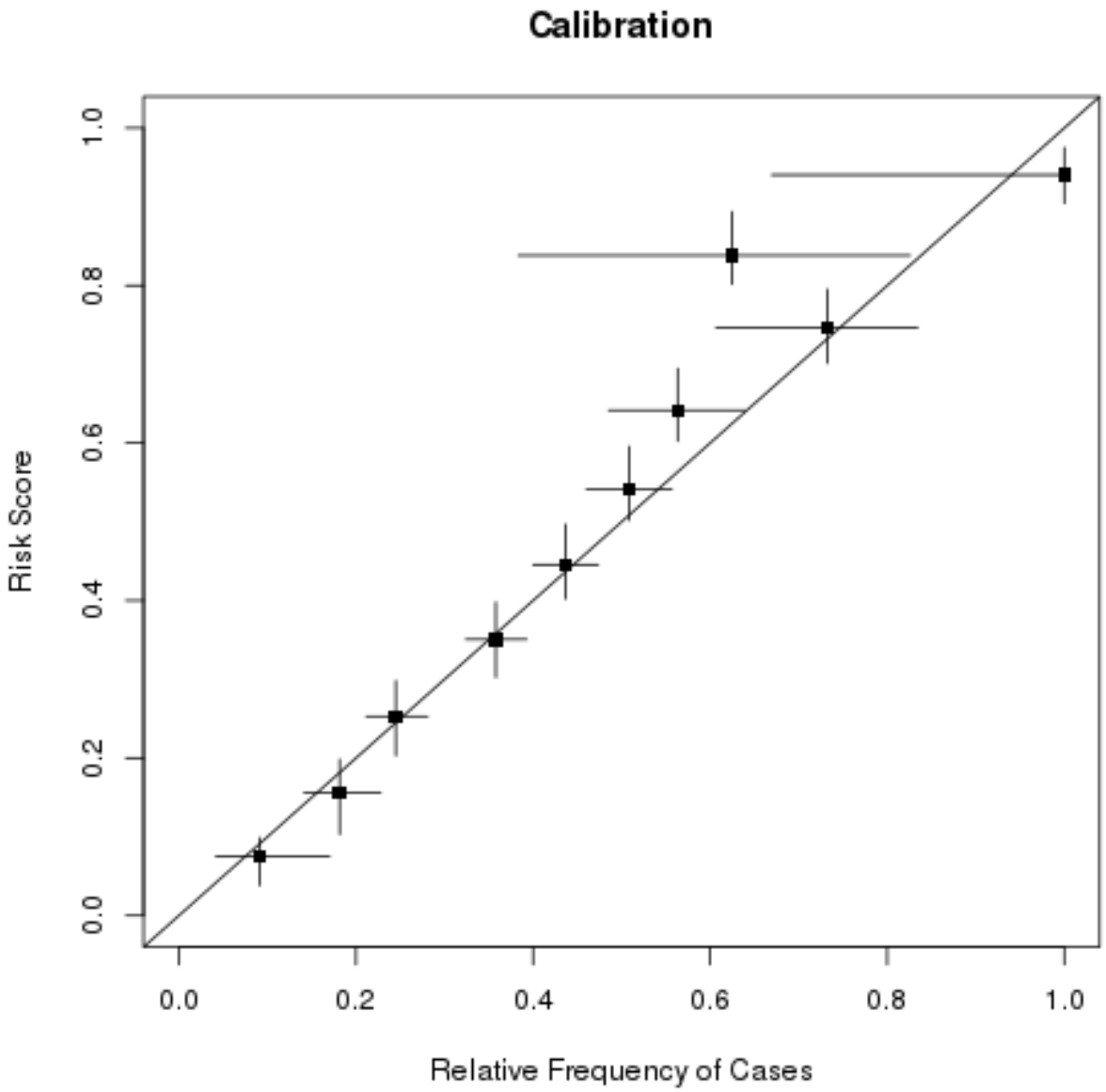

Calibration
Risk Score
Relative Frequency of Cases
0.0
0.2
0.4
0.6
0.8
1.0

Table 1. Description of 11 Case-Control Studies Included in the Invasive Epithelial Ovarian Cancer Relative Risk Prediction Model From the Ovarian Cancer Association Consortium (OCAC).

Study (Reference)	Study Name	Location	Period of Ascertainment	Age Range (Median) in Yrs	No. Controls	No. Cases	Response Rates[a]	
							Controls	Cases
CON (41)	Connecticut Ovarian Cancer Study	CT	1998-2003	34-81 (55)	466	318	61%	69%
DOV (14)	Diseases of the Ovary and their Evaluation	Western WA	2002-2009	35-74 (57)	1527	894	62%	74%
HAW (38)	Hawaii Ovarian Cancer Case-Control Study	HI, Southern CA	1993-2008	30-90 (57)	345	236	80%	78%
HOP (37)	Novel Risk Factors and Potential Early Detection Markers for Ovarian Cancer	Western PA, Northeast OH, Western NY	2003-2009	30-94 (57)	1561	570	68%	71%
MAY (36)	Mayo Clinic Ovarian Cancer Case-Control Study	IA, IL, MN, ND, SD, WI	2000-2010	30-92 (60)	842	533	58%	91%
NCO (42)	North Carolina Ovarian Cancer Study	NC	1999-2008	30-75 (57)	751	651	60%	67%
NEC (43)	New England Case-Control Study of Ovarian Cancer	NH, Eastern MA	1992-2003	30-78 (54)	1067	704	64%	71%
NJO (35)	New Jersey Ovarian Cancer Study	NJ	2002-2008	30-87 (60)	336	185	40%	47%
STA (39)	Genetic Epidemiology of Ovarian Cancer Study	San Francisco Bay Area, CA	1997-2001	30-65 (50)	330	276	75%	75%
UCI (45)	University of California Irvine Ovarian Study	Southern CA	1993-2005	30-86 (56)	505	318	80%	67%

USC (40, 44)	Los Angeles County Case-Control Studies of Ovarian Cancer	Los Angeles County, CA	1992-2002	30-85 (57)	1782	1108	72%	60%

Abbreviations: CA, California; CT, Connecticut; HI, Hawaii; IA, Iowa; IL, Illinois; MA, Massachusetts; MN, Minnesota; NC, North Carolina; ND, North Dakota; NH, New Hampshire; NJ, New Jersey; No, number; NY, New York; OH, Ohio; PA, Pennsylvania; SD, South Dakota; WA, Washington; WI, Wisconsin; Yrs, years.

[a]Response rates were calculated differently across studies; algorithms are available upon request.

Table 2. Frequency Distributions[a] of Risk Factors Included in the Invasive Epithelial Ovarian Cancer Relative Risk Prediction Model by Case-Control Status for the Training and Evaluation Sets.

Risk factors included in model	Training Set				Evaluation Set			
	Controls (n=7586)		Cases (n=4662)		Controls (n=1926)		Cases (n=1131)	
	N	(%)	N	(%)	N	(%)	N	(%)
Age at diagnosis/interview								
Mean (SD)	56.2	(11.6)	57.58	(10.9)	56.69	(11.7)	57.51	(10.9)
Age at menarche								
Mean (SD)	12.7	(1.6)	12.6	(1.5)	12.7	(1.5)	12.6	(1.5)
Missing age at menarche	63	(1)	95	(2)	19	(1)	28	(2)
Oral contraceptive use								
Ever Used	5341	(70)	2750	(60)	1350	(71)	682	(60)
Missing OC use	69	(1)	58	(1)	12	(1)	16	(1)
Mean months of OC use (SD)	74.7	(69.4)	58.3	(61.3)	76.3	(70.9)	59.1	(55.0)
Median months of OC use	57		36		58		48	
Missing months of OC use	89	(1)	79	(2)	19	(1)	21	(2)
Pregnancy History								
Mean number of full-term pregnancies (SD)	2.2	(1.5)	1.9	(1.6)	2.2	(1.6)	1.9	(1.5)
Missing number of full-term pregnancies	44	(1)	31	(1)	8	(<1)	10	(1)
Mean number of pregnancies (SD)	3.2	(1.7)	3.0	(1.7)	3.2	(1.7)	2.9	(1.6)
Missing number of pregnancies	45	(1)	31	(1)	8	(<1)	10	(1)
Mean number of non-full term pregnancies (SD)	0.65	(1.1)	0.52	(1.0)	0.60	(1.0)	0.53	(1.0)
Missing number of non-full term pregnancies	45	(1)	31	(1)	8	(<1)	10	(1)
Mean age at end of last pregnancy (SD)	30.5	(5.5)	29.5	(5.6)	30.7	(5.5)	29.8	(5.7)
Missing age at end of last pregnancy	638	(8)	413	(9)	162	(8)	94	(8)
Breastfeeding								
Ever breastfed	3250	(43)	1507	(32)	799	(41)	393	(35)
Missing breastfeeding status	1201	(16)	621	(13)	306	(16)	128	(11)
Mean months of breastfeeding (SD)	14.2	(16.3)	11.6	(15.8)	14.7	(15.8)	10.8	(12.7)
Missing breastfeeding duration	1203	(16)	623	(29)	306	(16)	128	(11)
Tubal ligation								
Had tubal ligation	1585	(21)	709	(15)	380	(20)	185	(16)
Missing tubal ligation	892	(12)	329	(7)	232	(12)	70	(6)

Endometriosis								
Had endometriosis	585	(8)	475	(10)	137	(7)	124	(11)
Missing endometriosis	354	(5)	367	(8)	78	(4)	93	(8)
Family history (1st degree relative)								
Breast cancer	1073	(14)	760	(16)	277	(14)	167	(15)
Missing breast cancer history	305	(4)	247	(5)	82	(4)	65	(6)
Ovarian cancer	202	(3)	239	(5)	55	(3)	53	(5)
Missing ovarian cancer history	397	(5)	284	(6)	99	(5)	78	(7)
Body mass index								
Mean BMI (SD)	26.44	(6.11)	26.82	(6.42)	26.50	(6.09)	26.47	(6.12)
Missing BMI	342	(5)	275	(6)	74	(4)	67	(6)
Aspirin use								
Irregular or non-user	3786	(50)	2349	(50)	975	(51)	572	(51)
Regular user of low-dose aspirin	186	(3)	64	(1)	46	(2)	19	(2)
Regular user of high-dose aspirin	247	(3)	103	(2)	49	(3)	38	(3)
Missing aspirin use	3367	(44)	2146	(46)	856	(44)	502	(44)
Menopausal status								
Post-menopausal	4818	(64)	3215	(69)	1247	(65)	774	(68)
Missing menopausal status	174	(2)	72	(2)	46	(2)	20	(2)
Hysterectomy								
Had hysterectomy[b]	1015	(13)	738	(16)	248	(13)	167	(15)
Missing hysterectomy	147	(2)	595	(13)	36	(2)	151	(13)
Menopausal hormone therapy								
Ever used MHT	2938	(39)	1907	(41)	749	(39)	477	(42)
Missing MHT use	108	(1)	139	(3)	30	(2)	42	(4)
Only used unopposed estrogen	833	(11)	642	(14)	206	(31)	152	(13)
Missing type of MHT	477	(6)	443	(10)	110	(11)	114	(10)
rs1243180[c]								
1 minor allele	2313	(41)	1512	(45)	631	(45)	342	(42)
2 minor alleles	523	(9)	368	(11)	140	(10)	86	(10)
rs2072590[c]								
1 minor allele	2414	(43)	1533	(45)	649	(46)	355	(43)
2 minor alleles	546	(10)	404	(12)	132	(9)	106	(13)
rs11782652[c]								
1 minor allele	734	(13)	476	(14)	163	(12)	125	(15)
2 minor alleles	25	(<1)	19	(1)	6	(<1)	5	(1)

rs10088218[c]								
1 minor allele	1306	(23)	689	(20)	348	(25)	185	(22)
2 minor alleles	105	(2)	43	(1)	21	(2)	9	(1)
rs757210[c]								
1 minor allele	2599	(46)	1567	(46)	662	(47)	379	(46)
2 minor alleles	762	(14)	525	(16)	180	(13)	123	(15)
rs9303542[c]								
1 minor allele	2219	(40)	1456	(43)	598	(43)	337	(41)
2 minor alleles	407	(7)	301	(9)	110	(8)	65	(8)
rs7651446[c]								
1 minor allele	527	(9)	423	(12)	121	(9)	117	(14)
2 minor alleles	15	(<1)	13	(<1)	7	(1)	9	(1)
rs3814113[c]								
1 minor allele	2421	(43)	1377	(41)	623	(44)	318	(39)
2 minor alleles	594	(11)	290	(9)	135	(10)	70	(8)
rs8170[c]								
1 minor allele	1735	(31)	1077	(32)	414	(30)	284	(34)
2 minor alleles	174	(3)	119	(4)	38	(3)	31	(4)
rs10069690[c]								
1 minor allele	2147	(39)	1350	(40)	523	(38)	322	(39)
2 minor alleles	351	(6)	234	(7)	101	(7)	58	(7)
rs12942666[c]								
1 minor allele	1719	(31)	1107	(33)	403	(29)	274	(33)
2 minor alleles	195	(3)	143	(4)	59	(4)	29	(4)

Abbreviations: BMI, body mass index; MHT, menopausal hormone therapy; N, number; OC, oral contraceptive; SD, standard deviation.

[a]Frequency distributions are based on non-missing data. Percent missing is based on the variable of interest and any upper level variable related to it. For example, women who are missing OC use status, and therefore duration of OC use, are combined with women who report ever using OCs but are missing duration of use to reach the number and percentage of women who are missing months of OC use.

[b]Women reporting hysterectomies more than one year prior to diagnosis (cases) or interview/reference date (controls) are considered to have had hysterectomy.

[c] Missing genotype data were approximately the same across the 11 SNPs: The percentage of participants missing genotype data was 26% (training set controls), 27%-28% (training set cases and evaluation set controls), and 27% (evaluation set cases).

Table 3. Risk Factors Included in the Invasive Epithelial Ovarian Cancer Relative Risk Prediction Model and Distributions and Covariates Used in Models to Impute Missing Values for Risk Factors with Missing Values.[a]

Risk factor	Covariates included in imputation model for Risk Factor | Distribution
SNP genotypes	Site | Multinomial-Dirichlet
Family history ovarian cancer	Site | Bernoulli
Family history breast cancer	Family history ovarian cancer, site | Bernoulli
Endometriosis	Cohort, age, site | Bernoulli
Menopausal status	Alcohol, smoking status, age, site | Bernoulli
Tubal ligation	Endometriosis, education, age, cohort, site | Bernoulli
Hysterectomy	Endometriosis, tubal ligation, family history breast cancer, family history ovarian cancer, age, cohort, site | Bernoulli
Height (BMI)	Site, cohort | Gaussian
Weight (BMI)	Site, cohort, height, age, smoking status, education | Gaussian
Aspirin use	Site, cohort, age, smoking status, BMI | Bernoulli
Ever used MHT	Menopausal status, hysterectomy, education, age, cohort, site | Bernoulli
Type of MHT	Ever used MHT, menopausal status, hysterectomy, education, age, cohort, site | Bernoulli
Age at menarche	Age, cohort, site | truncated Student t
Ever used OCs	Cohort, site | Bernoulli
Duration OC use	Ever used OCs, age, cohort, site | truncated Gaussian
Number of pregnancies	Hysterectomy, tubal ligation, ever used OCs, endometriosis, education, smoking, alcohol, age, cohort, site | PoissonNumber of Pregnancies, Hysterectomy, tubal ligation, ever used OCs, endometriosis, education, smoking, alcohol, age, cohort, site | Binomial
Number of full-term births	
Age at end of last pregnancy	Number of pregnancies, age at menarche, smoking status, education, age, cohort, site | truncated Gaussian

Ever breastfed	Number of pregnancies, smoking status, education, cohort, site | Bernoulli
Duration breastfeeding	Number of pregnancies, smoking status, education, age, cohort, site | truncated Gaussian

Abbreviations: BMI, body mass index; MHT, menopausal hormone therapy; OC, oral contraceptive; SNP, single nucleotide polymorphism.
[a]Left hand side variables (i.e., risk factors) may depend on any covariates given in the right hand column.

Table 4. Estimates of Log Odds Ratios (Medians) and 95% Bayesian Confidence Intervals for Risk Factors Included in the Invasive Epithelial Ovarian Cancer Relative Risk Prediction Model Containing 17 Confirmed SNPs, Stratified by Age (<50, ≥50) at Diagnosis (Cases) or Interview/Reference Age (Controls).[a]

Risk Factor	Age at Diagnosis/Interview <50		Age at Diagnosis/Interview ≥50	
	Median	**95% CI**	**Median**	**95% CI**
Age	**0.0308**	**0.0117, 0.0438**	-0.0067	-0.0205, 0.0014
High-dose Aspirin	0.05	-0.4624, 0.6254	-0.1223	-0.3517, 0.062
Low-dose Aspirin	-0.3338	-1.6847, 0.747	**-0.2982**	**-0.5838, -0.0262**
BMI	**0.0252**	**0.0148, 0.0381**	0.0023	-0.0059, 0.0087
Duration of Breastfeeding	-0.0079	-0.0166, 0.0001	**-0.0091**	**-0.0149, -0.0035**
Breastfeeding	**-0.3251**	**-0.5537, -0.0882**	-0.0342	-0.1658, 0.0889
Endometriosis	**0.5193**	**0.2967, 0.7637**	**0.2347**	**0.0645, 0.4095**
Family History Breast Cancer	**0.317**	**0.0885, 0.5534**	**0.1663**	**0.0537, 0.2902**
Family History Ovarian Cancer	**1.3687**	**0.9383, 1.7791**	**0.4949**	**0.2625, 0.7273**
Hysterectomy and No MHT	**-0.7656**	**-1.2045, -0.3448**	-0.0592	-0.2585, 0.1699
Age at End of Last Pregnancy	**-0.0148**	**-0.0289, -0.0024**	-0.005	-0.0108, 0.0017
Age at Menarche	**-0.0891**	**-0.1389, -0.0373**	0.0067	-0.0259, 0.0315
Menopausal Status	0.1161	-0.18, 0.3834	0.0955	-0.0744, 0.2697
MHT Estrogen without Hysterectomy	**1.5661**	**0.992, 1.8842**	-0.1107	-0.3277, 0.1101
MHT Estrogen and Hysterectomy	**-2.1774**	**-2.7231, -1.5081**	0.2408	-0.027, 0.4781
MHT Other without Hysterectomy	0.1682	-0.2312, 0.482	**-0.182**	**-0.3235, -0.0267**
MHT Other and Hysterectomy	1.2814	-0.1834, 2.5757	0.0166	-0.3454, 0.5927
Ever Used OCs	**-0.219**	**-0.4963, -0.0029**	-0.0069	-0.1703, 0.1463
Duration OC Use	**-0.1275**	**-0.1521, -0.1008**	**-0.0546**	**-0.0756, -0.0374**
Non-Full-Term Pregnancies	-0.1005	-0.2088, 0.0233	**-0.0719**	**-0.1144, -0.034**
Full-Term Births	**-0.1227**	**-0.203, -0.0463**	**-0.0644**	**-0.1188, -0.0166**
Tubal Ligation	**-0.4349**	**-0.6769, -0.2126**	**-0.2668**	**-0.4027, -0.1423**
rs1243180	0.1089	-0.0116, 0.2168	**0.1499**	**0.0806, 0.2232**
rs2072590	**0.1653**	**0.0695, 0.2806**	**0.1342**	**0.0629, 0.2034**
rs11782652	0.0686	-0.0858, 0.2117	0.0765	-0.037, 0.1985
rs10088218	**-0.1946**	**-0.3243, -0.0688**	**-0.1644**	**-0.2719, -0.0647**
rs757210	0.0275	-0.0711, 0.1192	**0.0757**	**0.0048, 0.1472**
rs9303542	**0.1151**	**0.003, 0.216**	**0.1857**	**0.1078, 0.2599**
rs7651446	**0.266**	**0.0877, 0.4144**	**0.2974**	**0.1702, 0.4162**

rs3814113	**-0.1142**	**-0.2172, -0.0052**	**-0.1719**	**-0.2483, -0.1062**
rs8170	0.0368	-0.0851, 0.1388	0.0771	-0.0028, 0.161
rs10069690	0.0236	-0.1049, 0.115	**0.1044**	**0.0332, 0.1843**
rs56318008	**0.1816**	**0.0705, 0.3095**	**0.1825**	**0.0862, 0.2661**
rs58722170	-0.028	-0.1337, 0.0807	0.0156	-0.0587, 0.0929
rs17329882	0.11	-0.0026, 0.2086	**0.1441**	**0.0749, 0.2237**
rs116133110	-0.0788	-0.1743, 0.0271	**-0.085**	**-0.1608, -0.0139**
rs635634	0.0644	-0.0627, 0.1807	0.071	-0.0135, 0.1492
chr17_29181220	-0.0946	-0.2029, 0.0192	**-0.1193**	**-0.1914, -0.0463**
rs183211	**0.1355**	**0.0323, 0.2447**	**0.0989**	**0.0318, 0.162**

Abbreviations: BMI, body mass index; CI, confidence interval; MHT, menopausal hormone therapy; N/A, not applicable; OC, oral contraceptive.
[a] Estimates and intervals are based on the training set only.

Table 5. Predictive power for relative risk prediction models for invasive epithelial ovarian cancer that include age, study site, 17 epidemiological risk factors, or 17 confirmed genetic susceptibility variants.

Age	Study Site	Epidemiological Risk Factors	SNPs	ROC AUC
Included	Included	Included	Included	0.664
Included	Included	Included	Not Included	0.649
Included	Included	Not Included	Included	0.601
Included	Included	Not Included	Not Included	0.563

Abbreviations: ROC AUC, receiver operating characteristic curve area under the curve; SNPs, single nucleotide polymorphisms.

Appendix 3: JAGS output for OC3 risk prediction model algorithm

Fit of Absolute Risk Model to OC3 Phase I 80% Training Set

ESI

October

14, 2015

1 JAGs Baseline Model for Phase I OC3 Data

1.1 Load Data

```
load("/proj/pooc3s/pooc30d/home/bl/oc3phaseI.data/oc3phaseI.RData")
```

1.2 JAGs Data Structure

```
  ## JAGS Data Structure: ##
  colnames(BL$mort)[colnames(BL$mort)=="96+"
  ]<-"96" minAge<-31
  ## Exclude women>80yo at BL and eval samples:
  x.train <- x[(x$train == 1)&(x$AgeAtBL<=80),]
  ## ******* Select 10% Sub-Samples: **********
  keep<-sample(1:nrow(x.train),size=floor(nrow(x.train)/10),replace=FALSE)
  x.train<-x.train[keep,]
  dim(x.train)
```

```
## [1] 32392    52 nsamp<-nrow(x.train)
```

```
  keep<-sample(1:nrow(OCAC),size=floor(nrow(OCAC)/5),replace=FALSE)
  ocac<-
  OCAC[keep,]
  nsamp0<-
  nrow(ocac)
  dim(ocac)
```

```
  ## [1]
3061         72
```

```
  ## ******* Structures & Variables: *********
  DM.study<-model.matrix(train~-1+factor(study),data=x.train)
  colnames(DM.study)<-
  substr(colnames(DM.study),14,nchar(colnames(DM.study)))
  DM.cohort<-model.matrix(train~-
  1+factor(cohort),data=x.train)
```

```
  colnames(DM.cohort)<-
  paste("c",substr(colnames(DM.cohort),15,nchar(colnames(DM.cohort))),sep="")
  nc<-ncol(DM.cohort)
  DM.cohort<-cbind(DM.cohort,rep(0,nrow(DM.cohort)),rep(0,nrow(DM.cohort)))
  colnames(DM.cohort)[(nc+1):(nc+2)]<-c("c1960","c1965")
  rm(nc)
  smoke.mat<-matrix(0,nrow=3,ncol=3)
smoke.mat[1,3]<-1 ## current smoker is index 1, cat 3
smoke.mat[2,1]<-1 ## never smoker is index 2, cat 1
smoke.mat[3,2]<-1 ## past smoker is index 3, cat 2
## Ask about these:
x.train$ocmos[(!is.na(x.train$ocever))&(x.train$ocever=="Ever")&(x.train$ocmos==0)]<-NA
x.train$ocmos[(!is.na(x.train$ocever))&(x.train$ocever=="Never")]<-NA
x.train$ul.ocdur<-((12*(x.train$AgeAtBL - 10))^(1/3))
x.train$ul.ocdur[x.train$AgeAtBL > 55]<-((12*(55 - 10))^(1/3))
## OCAC versions:
DM.study0<-model.matrix(case~-1+factor(site),data=ocac)
colnames(DM.study0)<-substr(colnames(DM.study0),13,15)
DM.cohort0<-model.matrix(case~-1+factor(cohort),data=ocac)
colnames(DM.cohort0)<-paste("c",substr(colnames(DM.cohort0),15,18),sep="")
ocac$ul.ocdur<-((12*(ocac$refage - 10))^(1/3))
ocac$ul.ocdur[ocac$refage > 55]<-((12*(55 - 10))^(1/3))
BLdat<-list(n.BLages=ncol(BL$mort),
            min.age=minAge,
            n.BLyears=nrow(BL$mort),
            BLages=as.numeric(colnames(BL$mort)),
            ## BLyears=as.numeric(rownames(BL£mort)),
            h.mort=BL$mort,
            h.all.a=BL$allinc.a,
            h.all.b=BL$allinc.b,
            h.ov.a=BL$ovinc.a,
            h.ov.b=BL$ovinc.b,
            bso.mu=bsoRatePars$bso.mu[1:3],
            bso.prec=solve(bsoRatePars$bso.var[1:3,1:3]),
            bsoLogRR.mu=(-2.910518), ## bsoRR.mean.ie2 (no brca+ )
            bsoLogRR.prec=6.997364,  ## bsoRR.sd.ie2^(-2)
            ## adjust following to col index in BL£mort, etc, structures
            a.a=floor(x.train$AgeAtBL - minAge + 1),
            a.f=floor(x.train$EventAge - minAge + 1),
            N=nsamp,
            Y=floor(x.train$BirthYr - 1900),
            Event=(1*(x.train$EventType>1)),
            Outcome=x.train$EventType,
            zero=matrix(0,nrow=nsamp,ncol=66),
            mu0=rep(0,100),
            prec1=diag(1,nrow=100,ncol=100),
            prec001=diag(0.01,nrow=100,ncol=100),
```

```
smoke=smoke.mat,
edu=diag(1,nrow=5,ncol=5),
precCCtoC=c(10000,4), ## sd=0.01 if "same"; sd=0.5 if "different"
n.rf=15,
p=ncol(DM.study),  ## ******* OC3 variables *******
X=DM.study, X.c=DM.cohort[,-
1], p.c=ncol(DM.cohort[,-1]),
ocever=x.train$oceverN,
ocdur=((x.train$ocmos)^(1/3)),
ul.ocdur=x.train$ul.ocduralc=
x.train$alcN,
fhbrca=x.train$fhbrcaN,
fhovca=x.train$fhovcaN,
edu.idx=x.train$educationN,
smoke.idx=x.train$smokeN,
menstat=x.train$menstatN,
mage=x.train$menarchage,
endo=x.train$endomN,
tlig=x.train$tligN,
  Outcome0=ocac$case,  ## ********* OCAC variables (end in 0) ******
  p0=ncol(DM.study0),
  N0=nsamp0,
  X0=DM.study0,
  X.c0=DM.cohort0[,-1],
  a.a0=floor(ocac$refage - minAge + 1),
  ocever0=ocac$oceverN,
  ocdur0=((ocac$ocmos)^(1/3)),
  ul.ocdur0=ocac$ul.ocdur,
  alc0=ocac$alcN,
  fhbrca0=ocac$fhbrcaN,
  fhovca0=ocac$fhovcaN,
  edu.idx0=ocac$educationN,
  smoke.idx0=ocac$smokeN,
  menstat0=ocac$menstatN,
  mage0=ocac$menarchage,
  tlig0=ocac$tligN,
  endo0=ocac$endomN)
```

1.3 Model

```
## Specification of Model in JAGS Language.  ##
BLmodel <- function() {
    ## Parameter structures and priors:
    for (y in 1:n.BLyears) {
        for (a in 1:n.BLages) {
            h.all[y, a] ~ dbeta(h.all.a[y, a], h.all.b[y, a]) %_%
                T(h.ov[y, a], 1)
```

```
            h.ov[y, a] ~ dbeta(h.ov.a[y, a], h.ov.b[y, a])
            h.othr[y, a] <- (h.all[y, a] - h.ov[y, a])
            h.ovB0[y, a] <- (h.ov[y, a]/(1 + ((bsoRR - 1) * bsoCDF[BLages[a]])))  ## BSO=0
            h.ovB1[y, a] <- (bsoRR * h.ovB0[y, a]) ## BSO=1
            h.eventB0[y, a] <- (h.mort[y, a] + h.othr[y, a] + h.ovB0[y,
                a])
            h.eventB1[y, a] <- (h.mort[y, a] + h.othr[y, a] + h.ovB1[y,
                a])
        }
    }
    ## Cumulative probabilities of surviving to age A=a BSO-free: Note
    ## that (bsoCDF(a)-bsoCDF(a-1)) = Pr(BSO(a)=1|M(a)=0,BSO(a-1)=0)
## Need: Pr(BSO(a.b)=1,BSO(a-1)=0,...,BSO(a0+1)=0|M(a)=0) and
## Pr(BSO(a.e)=0,BSO(a-1)=0,...,BSO(a0+1)=0|M(a)=0) where a.b=age
## at BSO and a.e is age at other event. a.b is set to a.e if no
## bso.  b = age at BL; a=age at BSO; a>=b b=1 <=> bl.age=30, b=66
## <=> bl.age=95
CPBSOFree[1] <- (1 - h.bso[1])
h.bso[1] <- (bsoCDF[1] - bsoRatePar[1] * pnorm(0, bsoRatePar[2],
    bsoRatePrec))
bsoCDF[1] <- bsoRatePar[1] * pnorm(1, bsoRatePar[2], bsoRatePrec)
for (a in 2:100) {
    CPBSOFree[a] <- CPBSOFree[a - 1] * (1 - h.bso[a])
    h.bso[a] <- (bsoCDF[a] - bsoCDF[a - 1])
    bsoCDF[a] <- bsoRatePar[1] * pnorm(a, bsoRatePar[2], bsoRatePrec)
}
bsoRatePrec <- bsoRatePar[3]^(-2)
bsoRatePar ~ dmnorm(bso.mu, bso.prec)
bsoRR ~ dlnorm(bsoLogRR.mu, bsoLogRR.prec) %_% I(, 1)
## Likelihood, OC3 Samples:
for (i in 1:N) {
    ## relative hazards
    rh.event[i] <- exp(inprod(alpha[], X[i, ]) + a[1] * ocever[i] +
        a[2] * fhbrca[i] + a[3] * fhovca[i] + inprod(a[4:7], edu[edu.idx[i],
        2:5]) + a[8] * alc[i] + inprod(a[9:10], smoke[smoke.idx[i],
        2:3]) + a[11] * menstat[i] + a[12] * ocdur[i] * ocever[i] +
        a[13] * mage[i] + a[14] * endo[i] + a[15] * tlig[i])
         rh.mort[i] <- exp(inprod(beta[], X[i, ]) + b[1] * ocever[i] +
        b[2] * fhbrca[i] + b[3] * fhovca[i] + inprod(b[4:7], edu[edu.idx[i],
        2:5]) + b[8] * alc[i] + inprod(b[9:10], smoke[smoke.idx[i],
        2:3]) + b[11] * menstat[i] + b[12] * ocdur[i] * ocever[i] +
        b[13] * mage[i] + b[14] * endo[i] + b[15] * tlig[i])
    rh.othr[i] <- exp(inprod(gamma[], X[i, ]) + g[1] * ocever[i] +
        g[2] * fhbrca[i] + g[3] * fhovca[i] + inprod(g[4:7], edu[edu.idx[i],
        2:5]) + g[8] * alc[i] + inprod(g[9:10], smoke[smoke.idx[i],
        2:3]) + g[11] * menstat[i] + g[12] * ocdur[i] * ocever[i] +
        g[13] * mage[i] + g[14] * endo[i] + g[15] * tlig[i])
    rh.ovca[i] <- exp(inprod(delta[], X[i, ]) + d[1] * ocever[i] +
        d[2] * fhbrca[i] + d[3] * fhovca[i] + inprod(d[4:7], edu[edu.idx[i],
```

```
    2:5]) + d[8] * alc[i] + inprod(d[9:10], smoke[smoke.idx[i],
    2:3]) + d[11] * menstat[i] + d[12] * ocdur[i] * ocever[i] +
    d[13] * mage[i] + d[14] * endo[i] + d[15] * tlig[i])
## 1=age30, 66=age95
for (a in a.a[i]:(a.f[i] - 1)) {
    pr.B0[i, a] <- (CPBSOFree[a + min.age - 1]/CPBSOFree[a.a[i] +
        min.age - 2])
    zero[i, a] ~ dpois((pr.B0[i, a] * h.eventB0[Y[i], a] +
        (1 - pr.B0[i, a]) * h.eventB1[Y[i], a]) * rh.event[i])
}
pr.atfuB0[i] <- (CPBSOFree[a.f[i] + min.age - 1]/CPBSOFree[a.a[i] +
    min.age - 2])
Event[i] ~ dpois((pr.atfuB0[i] * h.eventB0[Y[i], a.f[i]] +
    (1 - pr.atfuB0[i]) * h.eventB1[Y[i], a.f[i]]) * rh.event[i])
pi.event[i, 1] <- (1 - step(Event[i] - 0.5))
pi.event[i, 2] <- (h.mort[Y[i], a.f[i]] * rh.mort[i]) * step(Event[i] -
    0.5)
pi.event[i, 3] <- (h.othr[Y[i], a.f[i]] * rh.othr[i]) * step(Event[i] -
    0.5)
pi.event[i, 4] <- ((pr.atfuB0[i] * h.ovB0[Y[i], a.f[i]] +
    (1 - pr.atfuB0[i]) * h.ovB1[Y[i], a.f[i]]) * rh.ovca[i]) *
    step(Event[i] - 0.5) Outcome[i] ~
dcat(pi.event[i, ])
## Risk Factor Distributions:
ocever[i] ~ dbern(pi.ocever[i])
pi.ocever[i] <- ilogit(i.ocever + inprod(s.ocever[], X[i,
    2:p]) + inprod(c.ocever[], X.c[i, ]) + inprod(edu.ocever[1:4],
    edu[edu.idx[i], 2:5]))
ocdur[i] ~ dnorm(mu.ocdur[i], prec.ocdur) %_% T(0, ul.ocdur[i])
mu.ocdur[i] <- (i.ocdur + inprod(s.ocdur[], X[i, 2:p]) + inprod(c.ocdur[],
    X.c[i, ]) + age.ocdur * a.a[i] + inprod(edu.ocdur[1:4],
    edu[edu.idx[i], 2:5]))
fhbrca[i] ~ dbern(pi.fhbrca[i])
pi.fhbrca[i] <- ilogit(i.fhbrca + inprod(s.fhbrca[], X[i,
    2:p]))
fhovca[i] ~ dbern(pi.fhovca[i])
pi.fhovca[i] <- ilogit(i.fhovca + fhbrca.fhovca * fhbrca[i])
edu.idx[i] ~ dcat(pi.edu[i, 1:5])
pi.edu[i, 1] <- 1
pi.edu[i, 2] <- exp(i.edu[1] + inprod(s.edu[1, ], X[i, 2:p]) +
    inprod(c.edu[1, ], X.c[i, ]))
pi.edu[i, 3] <- exp(i.edu[2] + inprod(s.edu[2, ], X[i, 2:p]) +
    inprod(c.edu[2, ], X.c[i, ]))
pi.edu[i, 4] <- exp(i.edu[3] + inprod(s.edu[3, ], X[i, 2:p]) +
    inprod(c.edu[3, ], X.c[i, ]))
pi.edu[i, 5] <- exp(i.edu[4] + inprod(s.edu[4, ], X[i, 2:p]) +
    inprod(c.edu[4, ], X.c[i, ]))
alc[i] ~ dbern(pi.alc[i])
```

```
    pi.alc[i] <- ilogit(i.alc + inprod(s.alc[], X[i, 2:p]) + inprod(c.alc[],
        X.c[i, ]) + inprod(edu.alc[1:4], edu[edu.idx[i], 2:5]))
    smoke.idx[i] ~ dcat(pi.smoke[i, 1:3])
    pi.smoke[i, 1] <- 1
    pi.smoke[i, 2] <- exp(i.smoke[1] + inprod(s.smoke[1, ], X[i,
        2:p]) + inprod(c.smoke[1, ], X.c[i, ]) + alc.smoke[1] *
        alc[i] + inprod(edu.smoke[1, 1:4], edu[edu.idx[i], 2:5]))
    pi.smoke[i, 3] <- exp(i.smoke[2] + inprod(s.smoke[2, ], X[i,
        2:p]) + inprod(c.smoke[2, ], X.c[i, ]) + alc.smoke[2] *
        alc[i] + inprod(edu.smoke[2, 1:4], edu[edu.idx[i], 2:5]))
    menstat[i] ~ dbern(pi.meno[i])
    pi.meno[i] <- ilogit(i.meno + inprod(s.meno[], X[i, 2:p]) + age.meno *
        a.a[i] + alc.meno * alc[i] + inprod(smoke.meno[], smoke[smoke.idx[i],
        2:3]))
    mage[i] ~ dlnorm(mu.mage[i], prec.mage)
    mu.mage[i] <- i.mage + inprod(s.mage[], X[i, 2:p]) + inprod(c.mage[],
        X.c[i, ])
    endo[i] ~ dbern(pi.endo[i])
    pi.endo[i] <- ilogit(i.endo + inprod(s.endo[], X[i, 2:p]) +
        inprod(c.endo[], X.c[i, ]) + age.endo * a.a[i])
    tlig[i] ~ dbern(pi.tlig[i])
    pi.tlig[i] <- ilogit(i.tlig + inprod(s.tlig[], X[i, 2:p]) + inprod(c.tlig[],
        X.c[i, ]) + age.tlig * a.a[i] + inprod(edu.tlig[1:4], edu[edu.idx[i], 2:5]) +
        endo.tlig * endo[i])
}
## Likelihood, OCAC Samples:
for (i in 1:N0) {
    pi.case[i] <- ilogit(i.case0 + inprod(delta0[], X0[i, 2:p0]) + d0[1] *
        ocever0[i] + d0[2] * fhbrca0[i] + d0[3] * fhovca0[i] + inprod(d0[4:7],
        edu[edu.idx0[i], 2:5]) + d0[8] * alc0[i] + inprod(d0[9:10],
        smoke[smoke.idx0[i], 2:3]) + d0[11] * menstat0[i] + d0[12] *
        ocdur0[i] * ocever0[i] + d0[13] * mage0[i] + d0[14] * endo0[i] +
        d0[15] * tlig0[i])
    Outcome0[i] ~ dbern(pi.case[i])
    ## Risk Factor Distributions:
    ocever0[i] ~ dbern(pi.ocever0[i])
    pi.ocever0[i] <- ilogit(i.ocever + inprod(s.ocever0[], X0[i,
        2:p0]) + inprod(c.ocever[], X.c0[i, ]) + inprod(edu.ocever[1:4],
        edu[edu.idx0[i], 2:5]))
    ocdur0[i] ~ dnorm(mu.ocdur0[i], prec.ocdur) %_% T(0, ul.ocdur0[i])
    mu.ocdur0[i] <- (i.ocdur + inprod(s.ocdur0[], X0[i, 2:p0]) +
        inprod(c.ocdur[], X.c0[i, ]) + age.ocdur * a.a0[i] + inprod(edu.ocdur[1:4],
        edu[edu.idx0[i], 2:5]))
    fhbrca0[i] ~ dbern(pi.fhbrca0[i])
    pi.fhbrca0[i] <- ilogit(i.fhbrca + inprod(s.fhbrca0[], X0[i,
        2:p0]))
    fhovca0[i] ~ dbern(pi.fhovca0[i])
    pi.fhovca0[i] <- ilogit(i.fhovca + fhbrca.fhovca * fhbrca0[i])
```

```
    edu.idx0[i] ~ dcat(pi.edu0[i, 1:5])
    pi.edu0[i, 1] <- 1
    pi.edu0[i, 2] <- exp(i.edu[1] + inprod(s.edu0[1, ], X0[i,
        2:p0]) + inprod(c.edu[1, ], X.c0[i, ]))
    pi.edu0[i, 3] <- exp(i.edu[2] + inprod(s.edu0[2, ], X0[i,
        2:p0]) + inprod(c.edu[2, ], X.c0[i, ]))
    pi.edu0[i, 4] <- exp(i.edu[3] + inprod(s.edu0[3, ], X0[i,
        2:p0]) + inprod(c.edu[3, ], X.c0[i, ]))
    pi.edu0[i, 5] <- exp(i.edu[4] + inprod(s.edu0[4, ], X0[i,
        2:p0]) + inprod(c.edu[4, ], X.c0[i, ]))
    alc0[i] ~ dbern(pi.alc0[i])
    pi.alc0[i] <- ilogit(i.alc + inprod(s.alc0[], X0[i, 2:p0]) +
        inprod(c.alc[], X.c0[i, ]) + inprod(edu.alc[1:4], edu[edu.idx0[i],
        2:5]))
    smoke.idx0[i] ~ dcat(pi.smoke0[i, 1:3])
    pi.smoke0[i, 1] <- 1
    pi.smoke0[i, 2] <- exp(i.smoke[1] + inprod(s.smoke0[1, ],
        X0[i, 2:p0]) + inprod(c.smoke[1, ], X.c0[i, ]) + alc.smoke[1] *
        alc0[i] + inprod(edu.smoke[1, 1:4], edu[edu.idx0[i], 2:5]))
    pi.smoke0[i, 3] <- exp(i.smoke[2] + inprod(s.smoke0[2, ],
        X0[i, 2:p0]) + inprod(c.smoke[2, ], X.c0[i, ]) + alc.smoke[2] *
        alc0[i] + inprod(edu.smoke[2, 1:4], edu[edu.idx0[i], 2:5]))
    menstat0[i] ~ dbern(pi.meno0[i])
    pi.meno0[i] <- ilogit(i.meno + inprod(s.meno0[], X0[i, 2:p0]) + age.meno *
        a.a0[i] + alc.meno * alc0[i] + inprod(smoke.meno[], smoke[smoke.idx0[i],
        2:3]))
    mage0[i] ~ dlnorm(mu.mage0[i], prec.mage)
    mu.mage0[i] <- i.mage + inprod(s.mage0[], X0[i, 2:p0]) + inprod(c.mage[],
        X.c0[i, ])
    endo0[i] ~ dbern(pi.endo0[i])
    pi.endo0[i] <- ilogit(i.endo + inprod(s.endo0[], X0[i, 2:p0]) +
        inprod(c.endo[], X.c0[i, ]) + age.endo * a.a0[i])
    tlig0[i] ~ dbern(pi.tlig0[i])
    pi.tlig0[i] <- ilogit(i.tlig + inprod(s.tlig0[], X0[i, 2:p0]) + inprod(c.tlig[],
        X.c0[i, ]) + age.tlig * a.a0[i] + inprod(edu.tlig[1:4], edu[edu.idx0[i], 2:5]) +
        endo.tlig * endo0[i])
}
a ~ dmnorm(mu0[1:n.rf], prec1[1:n.rf, 1:n.rf]) b ~
dmnorm(mu0[1:n.rf], prec1[1:n.rf, 1:n.rf]) g ~
dmnorm(mu0[1:n.rf], prec1[1:n.rf, 1:n.rf]) for (i in
1:n.rf) {
    d[i] ~ dnorm(d0[i], precCCtoC[1 + CdiffCC[i]])
    CdiffCC[i] ~ dbern(pi.diff)
}
pi.diff ~ dbeta(1, 10)
d0 ~ dmnorm(mu0[1:n.rf], prec1[1:n.rf, 1:n.rf])
i.ocever ~ dnorm(0, 0.01) i.ocdur
~ dnorm(0, 0.01) i.fhbrca ~
```

```
dnorm(0, 0.01) i.fhovca ~ dnorm(0,
0.01)
i.edu ~ dmnorm(mu0[1:4], prec001[1:4, 1:4])
i.alc ~ dnorm(0, 0.01)
i.smoke ~ dmnorm(mu0[1:2], prec001[1:2, 1:2])
i.meno ~ dnorm(0, 0.01) i.mage ~
dnorm(0, 0.01) i.endo ~ dnorm(0,
0.01) i.tlig ~ dnorm(0, 0.01)
i.case0 ~ dnorm(0, 0.01)
fhbrca.fhovca ~ dnorm(0, 1)
s.ocever ~ dmnorm(mu0[1:(p - 1)], prec1[1:(p - 1), 1:(p - 1)])
s.ocever0 ~ dmnorm(mu0[1:(p0 - 1)], prec1[1:(p0 - 1), 1:(p0 -
    1)])
c.ocever ~ dmnorm(mu0[1:p.c], prec1[1:p.c, 1:p.c])
s.ocdur ~ dmnorm(mu0[1:(p - 1)], prec1[1:(p - 1), 1:(p - 1)]) s.ocdur0 ~
dmnorm(mu0[1:(p0 - 1)], prec1[1:(p0 - 1), 1:(p0 - 1)]) c.ocdur ~
dmnorm(mu0[1:p.c], prec1[1:p.c, 1:p.c])
age.ocdur ~ dnorm(0, 1)
s.alc ~ dmnorm(mu0[1:(p - 1)], prec1[1:(p - 1), 1:(p - 1)]) s.alc0 ~
dmnorm(mu0[1:(p0 - 1)], prec1[1:(p0 - 1), 1:(p0 - 1)]) c.alc ~
dmnorm(mu0[1:p.c], prec1[1:p.c, 1:p.c])
edu.ocever ~ dmnorm(mu0[1:4], prec1[1:4, 1:4])
edu.ocdur ~ dmnorm(mu0[1:4], prec1[1:4, 1:4])
edu.alc ~ dmnorm(mu0[1:4], prec1[1:4, 1:4])
s.fhbrca ~ dmnorm(mu0[1:(p - 1)], prec1[1:(p - 1), 1:(p - 1)])
s.fhbrca0 ~ dmnorm(mu0[1:(p0 - 1)], prec1[1:(p0 - 1), 1:(p0 -
    1)])
s.meno ~ dmnorm(mu0[1:(p - 1)], prec1[1:(p - 1), 1:(p - 1)]) s.meno0 ~
dmnorm(mu0[1:(p0 - 1)], prec1[1:(p0 - 1), 1:(p0 - 1)]) alc.meno ~ dnorm(0,
1)
age.meno ~ dnorm(0, 1)
smoke.meno ~ dmnorm(mu0[1:2], prec1[1:2, 1:2])
s.mage ~ dmnorm(mu0[1:(p - 1)], prec1[1:(p - 1), 1:(p - 1)]) s.mage0 ~
dmnorm(mu0[1:(p0 - 1)], prec1[1:(p0 - 1), 1:(p0 - 1)]) c.mage ~
dmnorm(mu0[1:p.c], prec1[1:p.c, 1:p.c])
s.endo ~ dmnorm(mu0[1:(p - 1)], prec1[1:(p - 1), 1:(p - 1)]) s.endo0 ~
dmnorm(mu0[1:(p0 - 1)], prec1[1:(p0 - 1), 1:(p0 - 1)]) c.endo ~
dmnorm(mu0[1:p.c], prec1[1:p.c, 1:p.c])
age.endo ~ dnorm(0, 1)
s.tlig ~ dmnorm(mu0[1:(p - 1)], prec1[1:(p - 1), 1:(p - 1)]) s.tlig0 ~
dmnorm(mu0[1:(p0 - 1)], prec1[1:(p0 - 1), 1:(p0 - 1)]) c.tlig ~
dmnorm(mu0[1:p.c], prec1[1:p.c, 1:p.c])
edu.tlig ~ dmnorm(mu0[1:4], prec1[1:4, 1:4])
age.tlig ~ dnorm(0, 1)
endo.tlig ~ dnorm(0, 1) for (i
in 1:4) {
    s.edu[i, 1:(p - 1)] ~ dmnorm(mu0[1:(p - 1)], prec1[1:(p -
        1), 1:(p - 1)])
```

```
    s.edu0[i, 1:(p0 - 1)] ~ dmnorm(mu0[1:(p0 - 1)], prec1[1:(p0 -
        1), 1:(p0 - 1)])
    c.edu[i, 1:p.c] ~ dmnorm(mu0[1:p.c], prec1[1:p.c, 1:p.c])
}
for (i in 1:2) {
    s.smoke[i, 1:(p - 1)] ~ dmnorm(mu0[1:(p - 1)], prec1[1:(p -
        1), 1:(p - 1)])
    s.smoke0[i, 1:(p0 - 1)] ~ dmnorm(mu0[1:(p0 - 1)], prec1[1:(p0 -
        1), 1:(p0 - 1)])
    c.smoke[i, 1:p.c] ~ dmnorm(mu0[1:p.c], prec1[1:p.c, 1:p.c])
    edu.smoke[i, 1:4] ~ dmnorm(mu0[1:4], prec1[1:4, 1:4])
}
alc.smoke ~ dmnorm(mu0[1:2], prec1[1:2, 1:2])
for (i in 1:p) {
    alpha[i] ~ dnorm(0, prec.event) beta[i]
    ~ dnorm(0, prec.mort) gamma[i] ~
    dnorm(0, prec.othr) delta[i] ~ dnorm(0,
    prec.ovca)
}
for (i in 1:(p0 - 1)) {
    delta0[i] ~ dnorm(0, prec.ovca)
}
prec.ocdur    <-    pow(sd.ocdur,    -2)
prec.event    <-    pow(sd.event,    -2)
prec.mort <- pow(sd.mort, -2)

    prec.othr   <-   pow(sd.othr,   -2)
    prec.ovca   <-   pow(sd.ovca,   -2)
    prec.mage   <-   pow(sd.ovca,   -2)
    sd.ocdur ~ dexp(1)
    sd.event ~ dexp(1)
    sd.mort ~ dexp(1)
    sd.othr ~ dexp(1)
    sd.ovca ~ dexp(1)
    sd.mage ~ dexp(1)
}
```

1.4 Initial Values

```
#################### Starting Values: #
BLinits <- function() {
    alc.init <- rep(1, length(BLdat$alc))
    alc.init[!(is.na(BLdat$alc))] <- NA alc.init0
    <- rep(1, length(BLdat$alc0))
    alc.init0[!(is.na(BLdat$alc0))] <- NA
    ocever.init <- rep(0, length(BLdat$ocever))
    ocever.init[!(is.na(BLdat$ocever))] <- NA
    ocever.init0 <- rep(0, length(BLdat$ocever0))
    ocever.init0[!(is.na(BLdat$ocever0))] <- NA
```

```
fhovca.init <- rep(0, length(BLdat$fhovca))
fhovca.init[!(is.na(BLdat$fhovca))] <- NA
fhovca.init0 <- rep(0, length(BLdat$fhovca0))
fhovca.init0[!(is.na(BLdat$fhovca0))] <- NA
fhbrca.init0 <- rep(0, length(BLdat$fhbrca0))
fhbrca.init0[!(is.na(BLdat$fhbrca0))] <- NA
smoke.init <- rep(NA, length(BLdat$smoke.idx))
smoke.init[is.na(BLdat$smoke.idx)] <- 1 smoke.init0
<- rep(NA, length(BLdat$smoke.idx0))
smoke.init0[is.na(BLdat$smoke.idx0)] <- 1 edu.init
<- rep(NA, length(BLdat$edu.idx))
edu.init[is.na(BLdat$edu.idx)] <- 1
edu.init0 <- rep(NA, length(BLdat$edu.idx0))
edu.init0[is.na(BLdat$edu.idx0)] <- 1 ocdur.init
<- rep(NA, length(BLdat$ocdur))
ocdur.init[is.na(BLdat$ocdur)] <- 2 ocdur.init0
<- rep(NA, length(BLdat$ocdur0))
ocdur.init0[is.na(BLdat$ocdur0)] <- 2 meno.init
<- rep(NA, length(BLdat$menstat))
meno.init[is.na(BLdat$menstat)] <- 1 meno.init0
<- rep(NA, length(BLdat$menstat0))
meno.init0[is.na(BLdat$menstat0)] <- 1 mage.init
<- rep(NA, length(BLdat$mage))
mage.init[is.na(BLdat$mage)] <- 13
mage.init0 <- rep(NA, length(BLdat$mage0))
mage.init0[is.na(BLdat$mage0)] <- 13
endo.init <- rep(0, length(BLdat$endo))
endo.init[!(is.na(BLdat$endo))] <- NA
endo.init0 <- rep(0, length(BLdat$endo0))
endo.init0[!(is.na(BLdat$endo0))] <- NA
tlig.init <- rep(0, length(BLdat$tlig))
tlig.init[!(is.na(BLdat$tlig))] <- NA
tlig.init0 <- rep(0, length(BLdat$tlig0))
tlig.init0[!(is.na(BLdat$tlig0))] <- NA
return(list(h.all = BL$allinc, h.ov = BL$ovinc, alpha = rep(0,
    ncol(DM.study)), beta = rep(0, ncol(DM.study)), gamma = rep(0,
    ncol(DM.study)), delta = rep(0, ncol(DM.study)), delta0 = rep(0,
    ncol(DM.study0) - 1), sd.ocdur = 1, sd.event = 0.1, sd.mort = 0.1,
    sd.othr = 0.1, sd.ovca = 0.1, bsoRatePar = bsoRatePars$bso.mu[1:3],
    bsoRR = 0.05, a = rep(0, BLdat$n.rf), b = rep(0, BLdat$n.rf),
    g = rep(0, BLdat$n.rf), d = rep(0, BLdat$n.rf), CdiffCC = rep(0,
        BLdat$n.rf), d0 = rep(0, BLdat$n.rf), i.ocever = 0.5,
    s.ocever = rep(0, (BLdat$p - 1)), c.ocever = rep(0, BLdat$p.c),
    edu.ocever = rep(0, 4), s.ocever0 = rep(0, (BLdat$p0 - 1)), i.ocdur
    = 3, s.ocdur = rep(0, (BLdat$p - 1)), c.ocdur = rep(0,
        BLdat$p.c), edu.ocdur = rep(0, 4), age.ocdur = 0.1, s.ocdur0 = rep(0,
        (BLdat$p0 - 1)), i.edu = rep(1, 4), s.edu = matrix(0,
        nrow = 4, ncol = (BLdat$p - 1)), c.edu = matrix(0, nrow = 4,
```

```
        ncol = BLdat$p.c), s.edu0 = matrix(0, nrow = 4, ncol = (BLdat$p0 -
        1)), i.smoke = rep(1, 2), s.smoke = matrix(0, nrow = 2,
        ncol = (BLdat$p - 1)), c.smoke = matrix(0, nrow = 2, ncol = BLdat$p.c),
    s.smoke0 = matrix(0, nrow = 2, ncol = (BLdat$p0 - 1)), edu.smoke = matrix(0,
        nrow = 2, ncol = 4), alc.smoke = rep(0, 2), i.alc = 1,
    s.alc = rep(0, (BLdat$p - 1)), c.alc = rep(0, BLdat$p.c),
    edu.alc = rep(0, 4), s.alc0 = rep(0, (BLdat$p0 - 1)), i.fhbrca = (-1),
    s.fhbrca = rep(0, (BLdat$p - 1)), s.fhbrca0 = rep(0, (BLdat$p0 -
        1)), i.fhovca = (-2), fhbrca.fhovca = 1, i.meno = 1, age.meno = 0,
    alc.meno = 0, smoke.meno = rep(0, 2), s.meno = rep(0, (BLdat$p -
        1)), s.meno0 = rep(0, (BLdat$p0 - 1)), i.mage = 2.5, s.mage = rep(0,
        (BLdat$p - 1)), s.mage0 = rep(0, (BLdat$p0 - 1)), c.mage = rep(0,
        BLdat$p.c), sd.mage = 0.12, i.endo = 0.15, s.endo = rep(0,
        (BLdat$p - 1)), c.endo = rep(0, BLdat$p.c), age.endo = 0, s.endo0
    = rep(0, (BLdat$p0 - 1)), i.tlig = 0.15, s.tlig = rep(0,
        (BLdat$p - 1)), c.tlig = rep(0, BLdat$p.c), age.tlig = 0, s.tlig0
    = rep(0, (BLdat$p0 - 1)), endo.tlig = 0, edu.tlig = rep(0,
        4), mage = mage.init, mage0 = mage.init0, endo = endo.init,
    endo0 = endo.init0, tlig = tlig.init, tlig0 = tlig.init0,
    ocever = ocever.init, ocever0 = ocever.init0, ocdur = ocdur.init,
    ocdur0 = ocdur.init0, fhovca = fhovca.init, fhovca0 = fhovca.init0,
    fhbrca0 = fhbrca.init0, alc = alc.init, alc0 = alc.init0,
    smoke.idx = smoke.init, smoke.idx0 = smoke.init0, edu.idx = edu.init,
    edu.idx0 = edu.init0))
}


library(coda)
library(rjags)
library(R2WinBUGS)
library(R2jags) fun.model.file
<- "BLmodel"
write.model(BLmodel, fun.model.file)
BLparameters <- c("bsoRatePar", "bsoRR", "alpha", "beta", "gamma", "delta",
    "delta0", "sd.event", "sd.mort", "sd.othr", "sd.ovca", "sd.ocdur",
    "i.fhovca", "fhbrca.fhovca", "i.ocever", "s.ocever", "c.ocever",
    "edu.ocever", "s.ocever0", "i.fhbrca", "s.fhbrca", "s.fhbrca0", "i.edu",
    "s.edu", "c.edu", "s.edu0", "i.alc", "s.alc", "c.alc", "edu.alc",
    "s.alc0", "i.smoke", "s.smoke", "c.smoke", "alc.smoke", "edu.smoke",
    "s.smoke0", "i.meno", "s.meno", "alc.meno", "age.meno", "smoke.meno",
    "s.meno0", "i.ocdur", "s.ocdur", "c.ocdur", "age.ocdur", "edu.ocdur",
    "s.ocdur0", "i.mage", "s.mage", "s.mage0", "c.mage", "sd.mage", "i.endo",
    "s.endo", "c.endo", "age.endo", "s.endo0", "i.tlig", "s.tlig", "c.tlig",
    "age.tlig", "s.tlig0", "edu.tlig", "endo.tlig", "CdiffCC", "pi.diff",
    "a", "b", "g",
    "d", "d0")
```

2 JAGS Parameter Estimation

```
system.time(BLjags <- jags(data = BLdat, inits = BLinits, parameters = BLparameters,
    fun.model.file, progress.bar = "none", n.chains = 1, n.iter = 3500,
    n.burnin = 1000, n.thin = 5, DIC = F))
```

```
## Compiling model graph
##    Resolving undeclared variables
##    Allocating nodes
##    Graph Size: 5011276
##
## Initializing model
##       user     system    elapsed
## 172837.351    15.701 173010.576
```

```
## ---- Run Times: ------------------------------------------
## nsamp n.iter time vars 32.4K 3.5K 47198=13.1hrs Study 32.4K 3.5K
## 81707=22.7hrs Study+FH.bc+FH.ov+ocever 32.4K 3.5K 66087=18.4hrs
## Study+FH.bc+FH.ov+ocever+edu 32.4K 3.5K 69081=19.2hrs
## Study+FH.bc+FH.ov+ocever+edu+smoke+alc 32.4K 3.5K 83680=23.2hrs
## Study+FH.bc+FH.ov+ocever+edu+smoke+alc+meno 32.4K 3.5K
## 115425=32.1hrs Study+FH.bc+FH.ov+ocever+edu+smoke+alc+meno+ocdur
## ---- ++(20% OCAC) ------ 32.4K 3.5K 145760=40.5hrs
## Study+FH.bc+FH.ov+ocever+edu+smoke+alc+meno+ocdur 32.4K 3.5K
## 128950=35.8hrs
## Study+FH.bc+FH.ov+ocever+edu+smoke+alc+meno+ocdur+mage 32.4K
## 3.5K 134437=37.3hrs
## Study+FH.bc+FH.ov+ocever+edu+smoke+alc+meno+ocdur+mage+endo
## 32.4K 3.5K hrs
## Study+FH.bc+FH.ov+ocever+edu+smoke+alc+meno+ocdur+mage+endo+tlig
```

```
dim(simMatrix <- BLjags$BUGSoutput$sims.matrix)
```

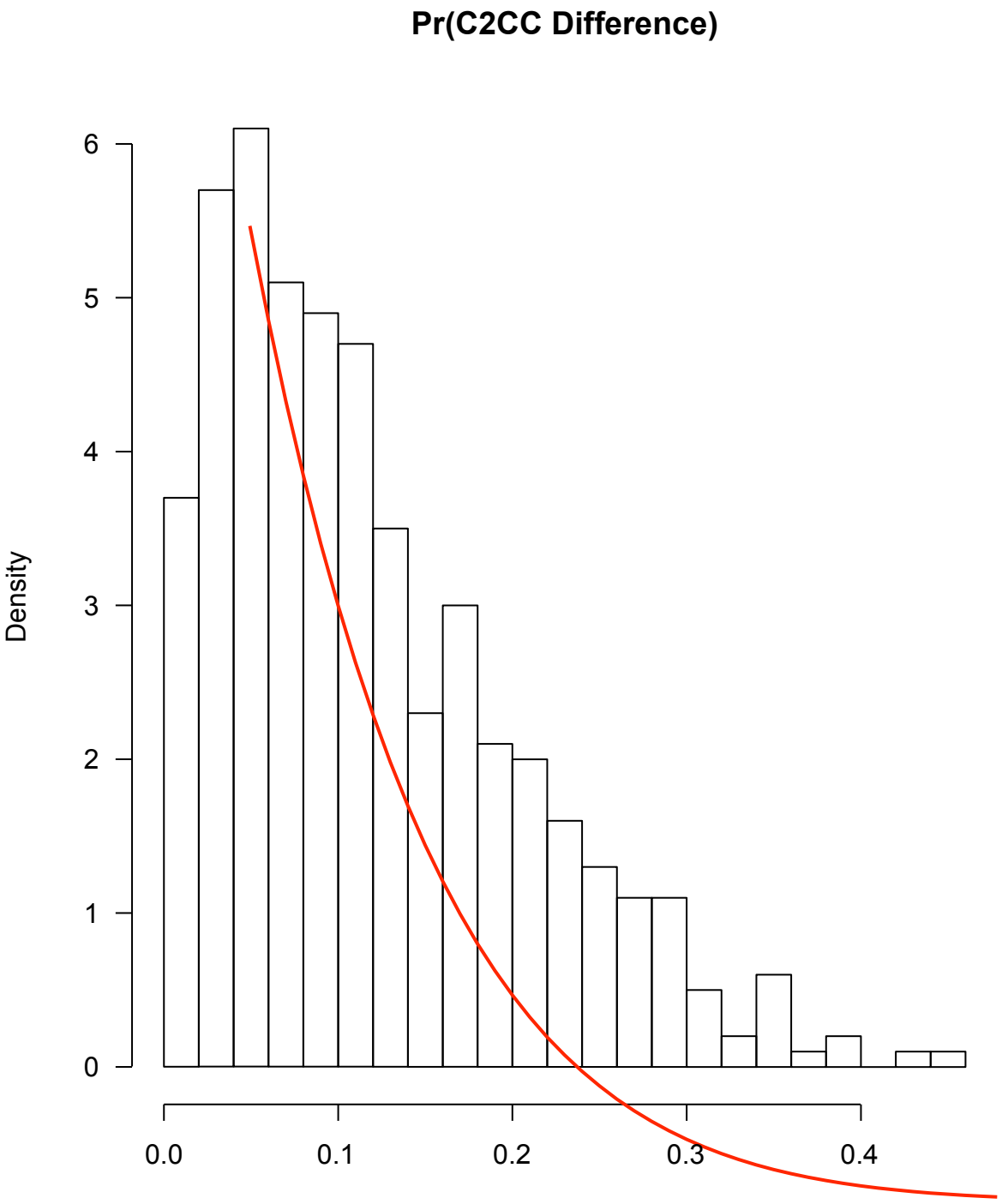
```
## [1] 500 521
```

```
## Risk Factor Labels
rf.names <- c("ocever", "fhbrca", "fhovca", "edu.col", "edu.grad", "edu.hs",
    "edu.lths", "alc", "smoke.past", "smoke.now", "meno.stat", "ocdur",
    "mage", "endo", "tlig")
## Pr(Case Control -- Cohort Effect Difference):
prDiff <- apply(simMatrix[, substr(colnames(simMatrix), 1, 7) == "CdiffCC"],
    2, mean)
names(prDiff) <- rf.names
prDiff
```

```
##     ocever     fhbrca     fhovca    edu.col   edu.grad     edu.hs
##      0.076      0.226      0.034      0.072      0.024      0.054
##   edu.lths        alc smoke.past  smoke.now  meno.stat      ocdur
##      0.696      0.040      0.124      0.168      0.204      0.026
```

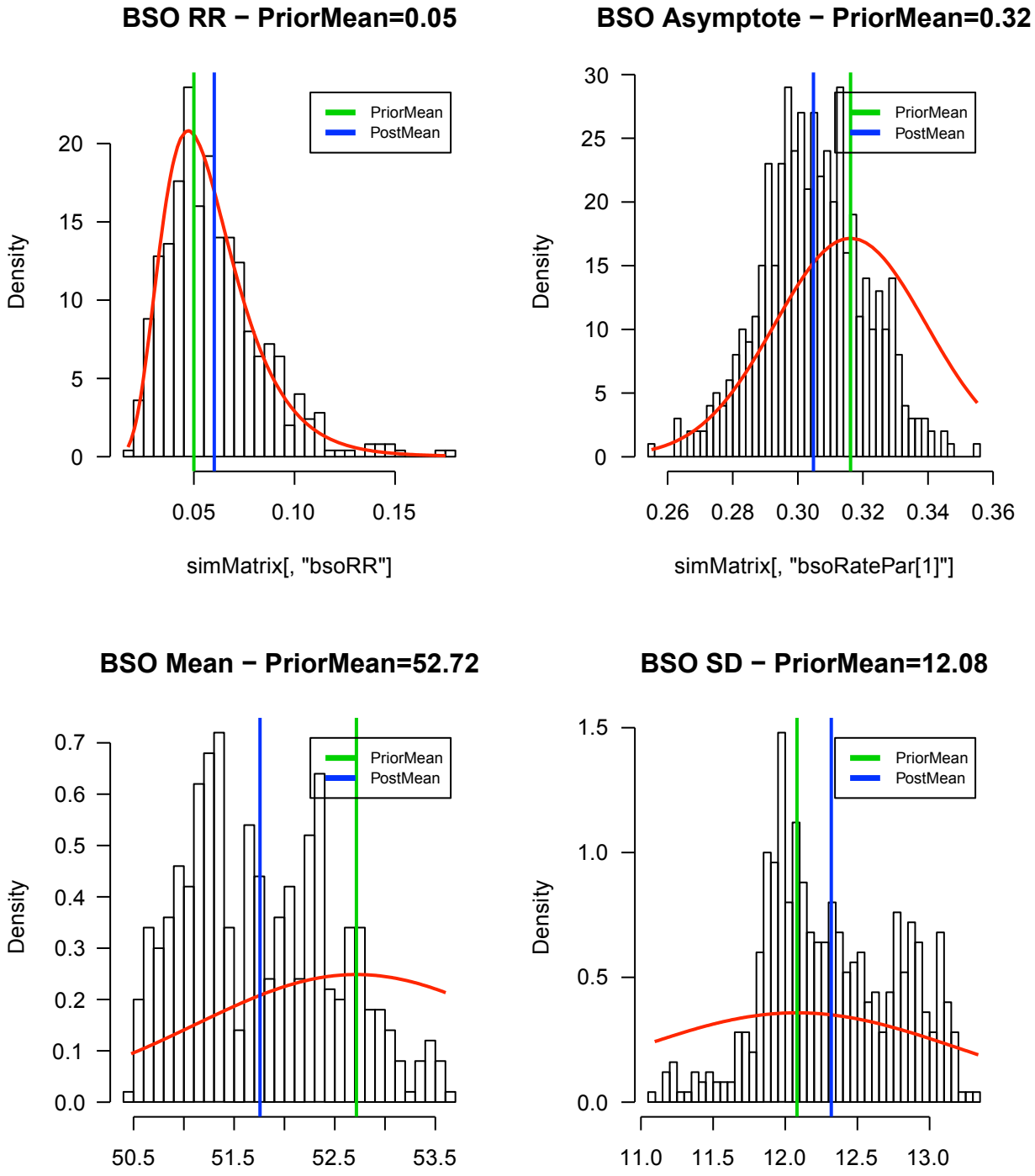
```
##       mage       endo       tlig
##      0.108      0.122      0.062
```

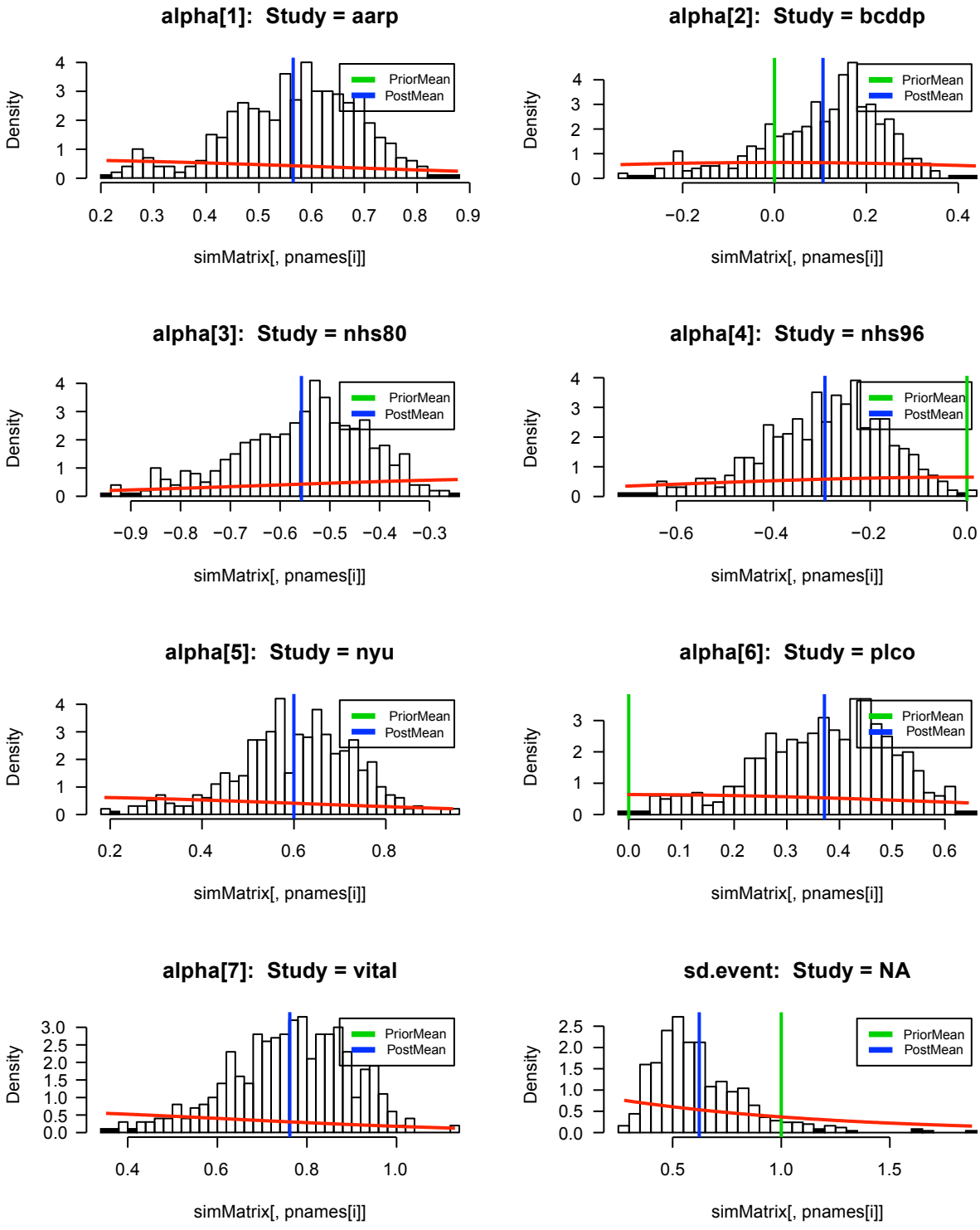
```
hist(simMatrix[, "pi.diff"], prob = TRUE, las = 1, main = "Pr(C2CC Difference)",
    xlab = "Pr(C2CC Difference)", nclass = 20)
grid <- seq(0, 1, by = 0.01)
lines(grid, dbeta(grid, 1, 10), col = 2, lwd = 2)
```



Pr(C2CC Difference)

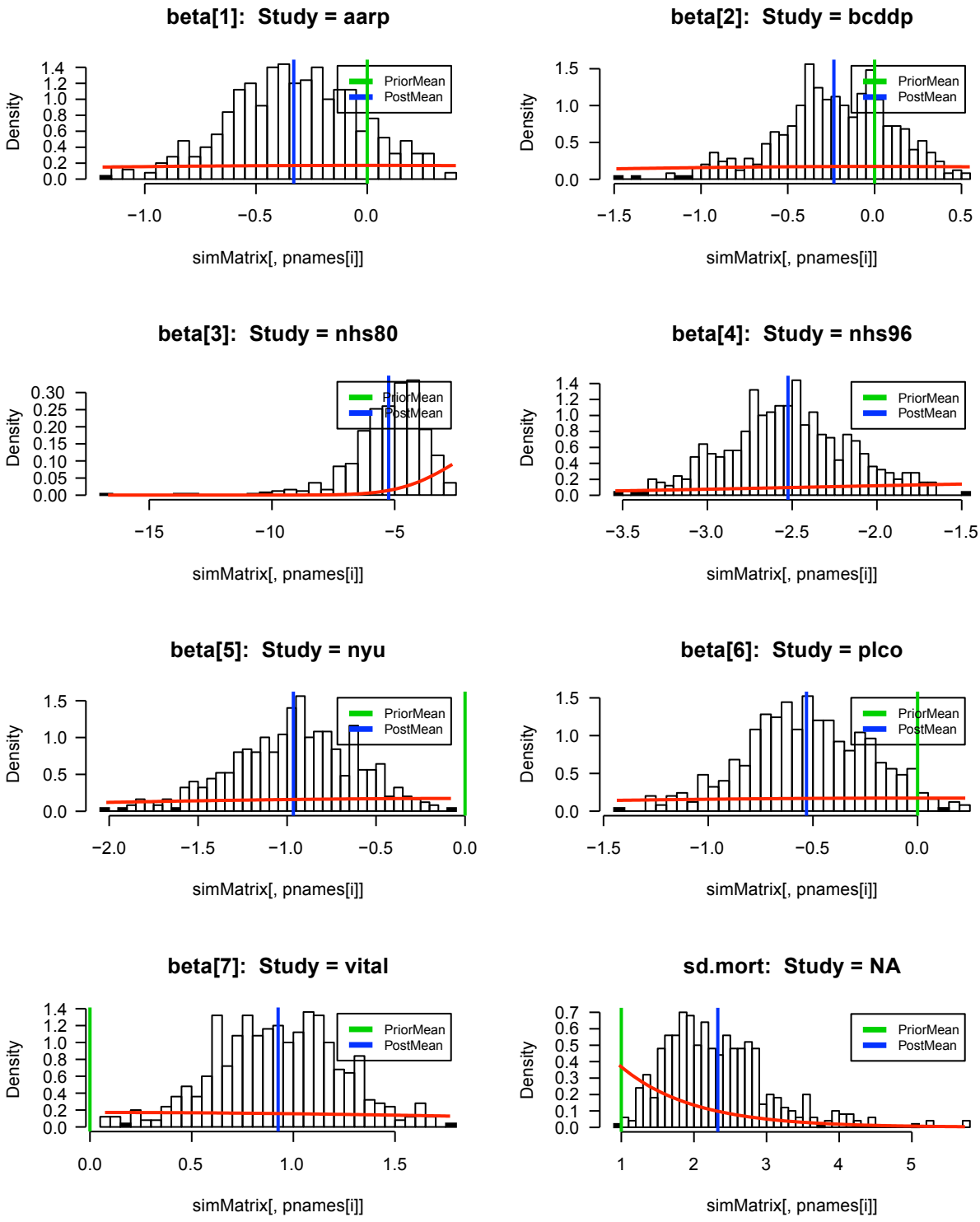
```
## Marginal Posterior Histograms
source("Functions.R")
margPostHist(par.set = "bso", layout = c(2, 2))
```





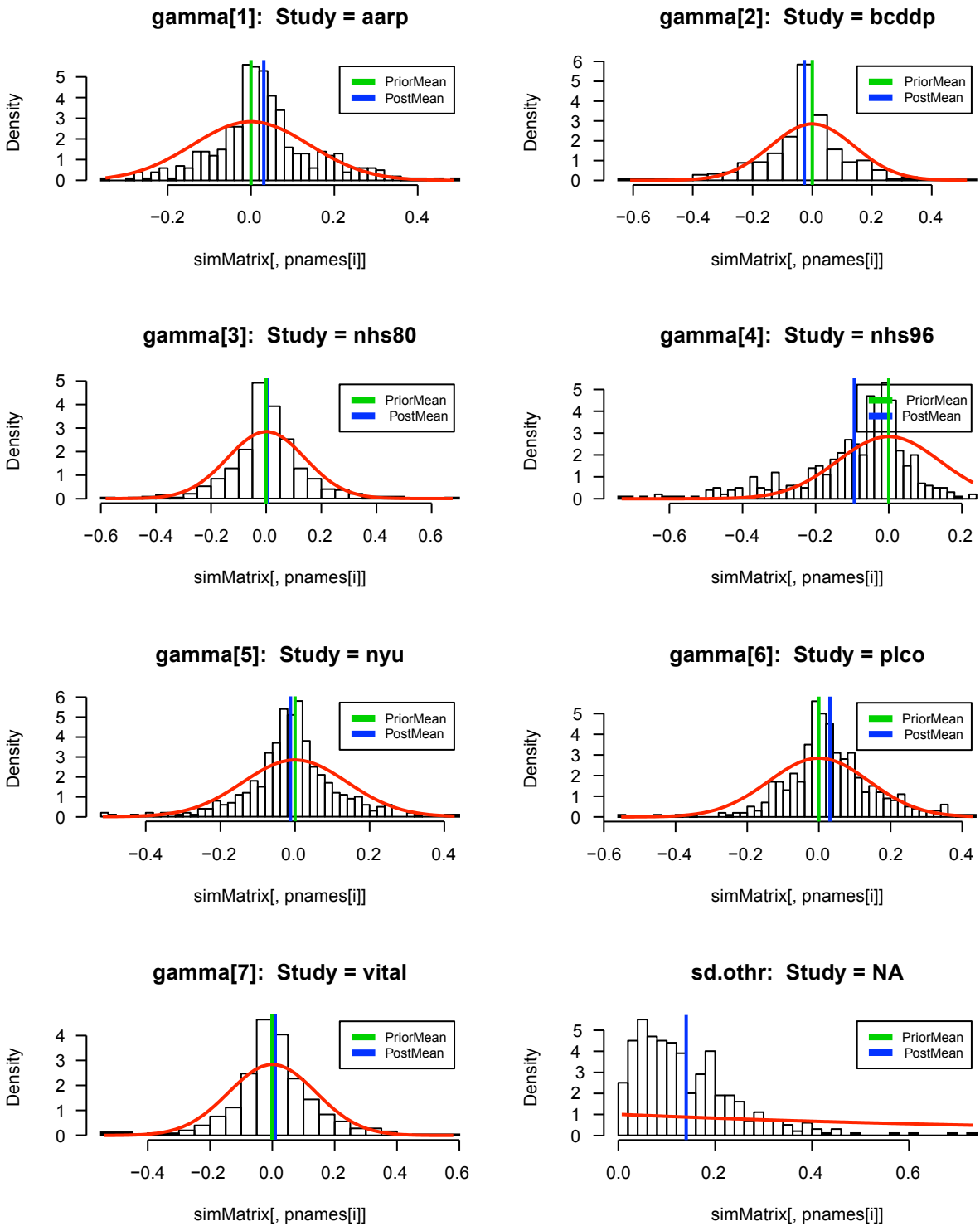

```
## NULL
```

```
margPostHist(par.set = "mort", layout = c(4, 2))
```



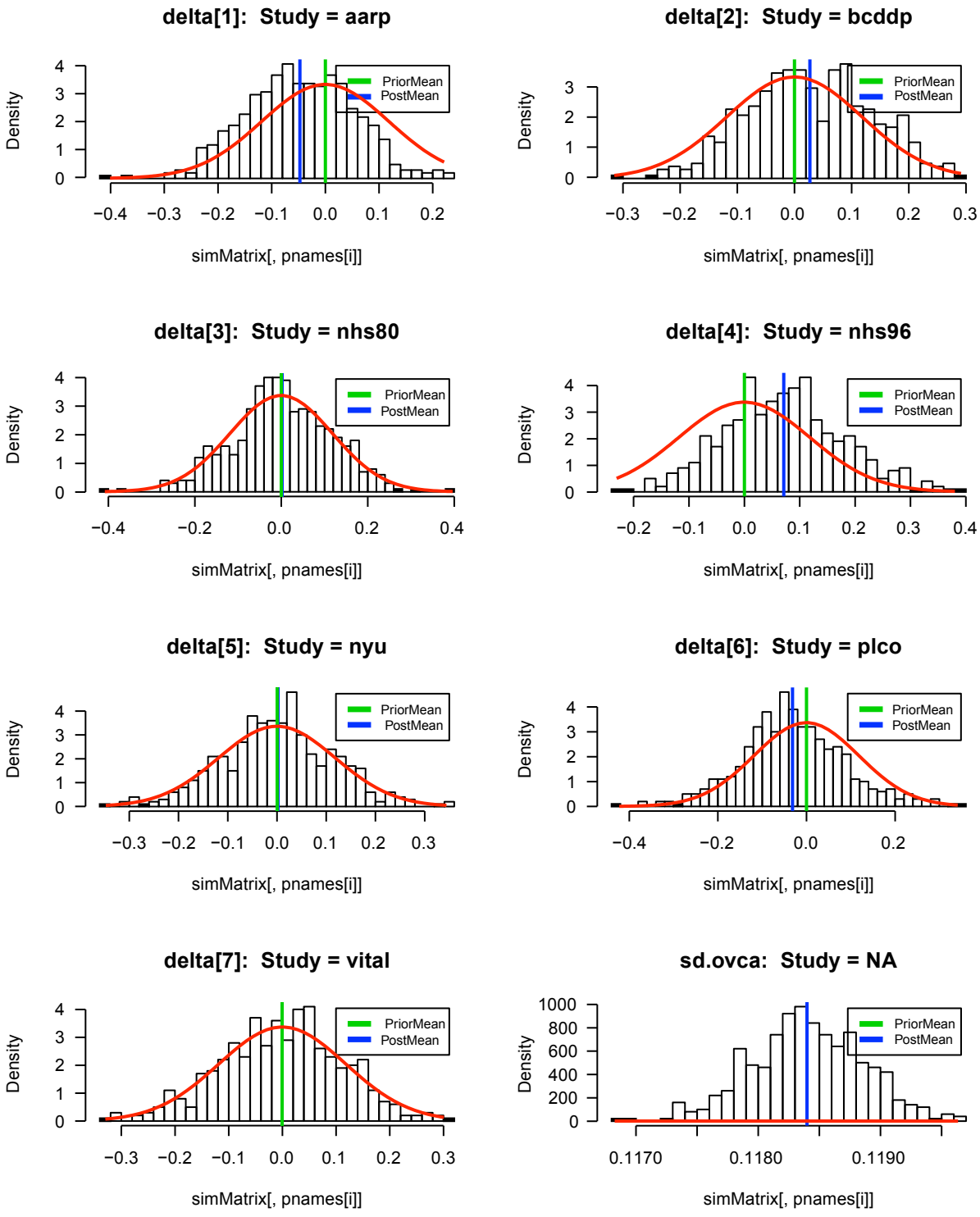

```
## NULL
```

```
margPostHist(par.set = "othr", layout = c(4, 2))
```



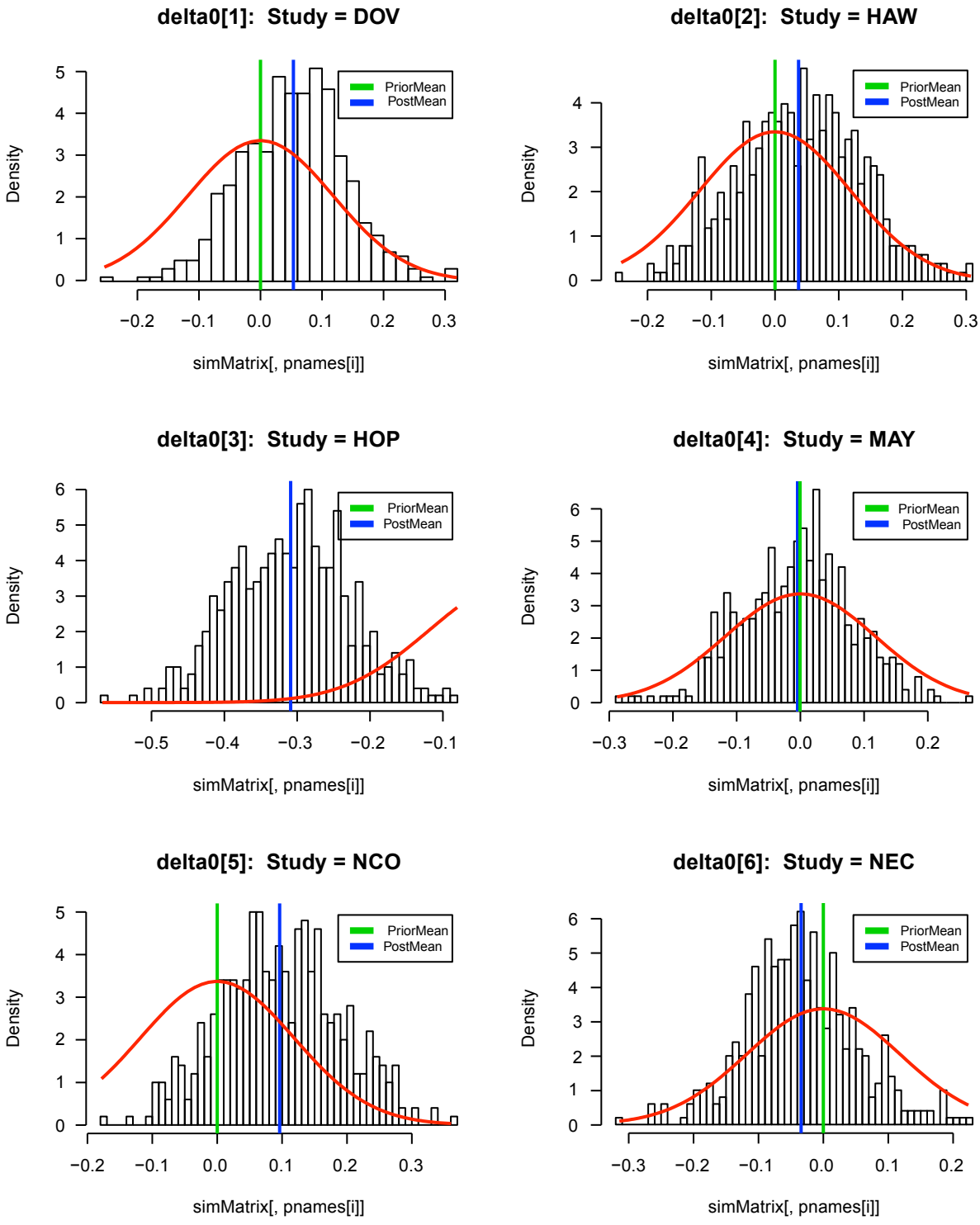

```
## NULL
```

```
margPostHist(par.set = "ovca.co", layout = c(4, 2))
```



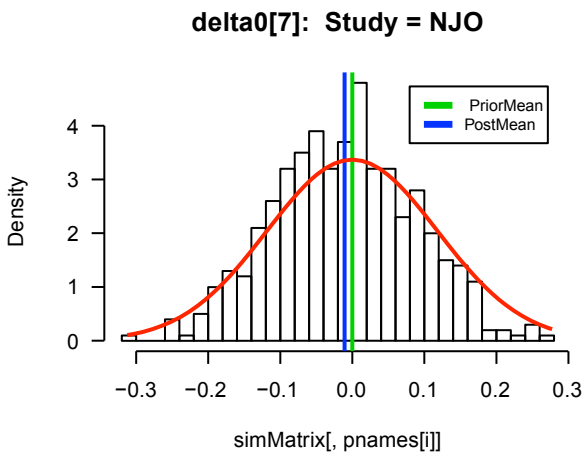

```
## NULL
```

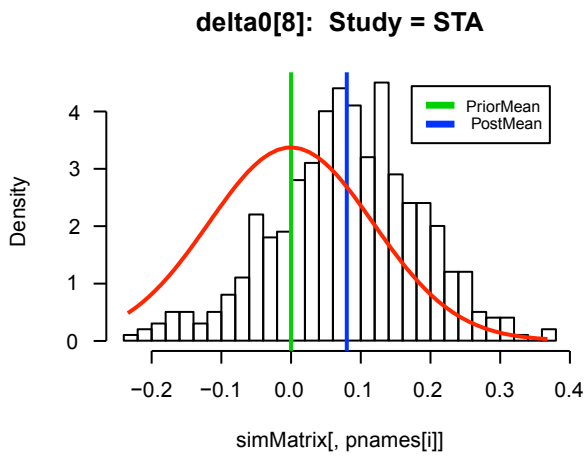
```
margPostHist(par.set = "ovca.cc", layout = c(3, 2))
```

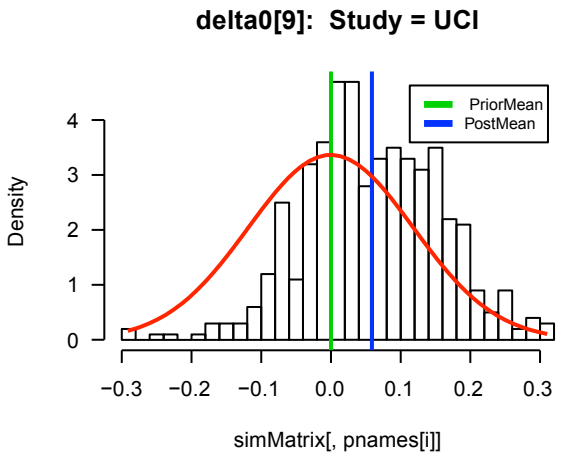


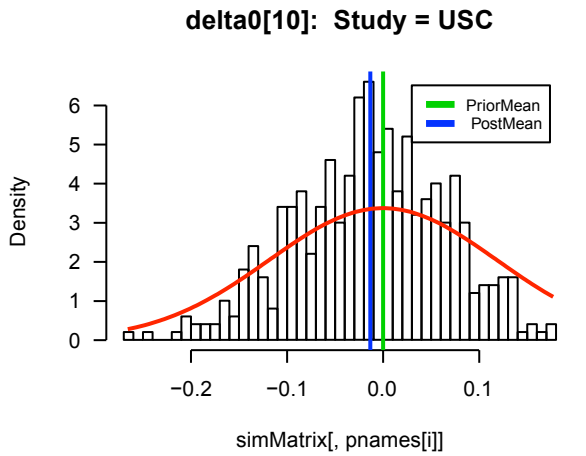

```
## NULL
```

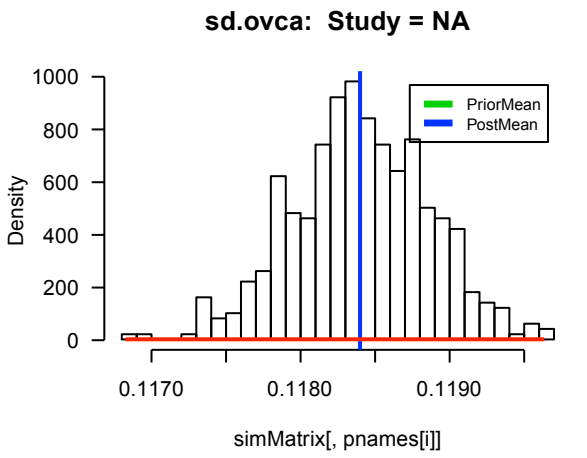
```
rfplots <- paste("rf", 1:BLdat$n.rf, sep = "")
```

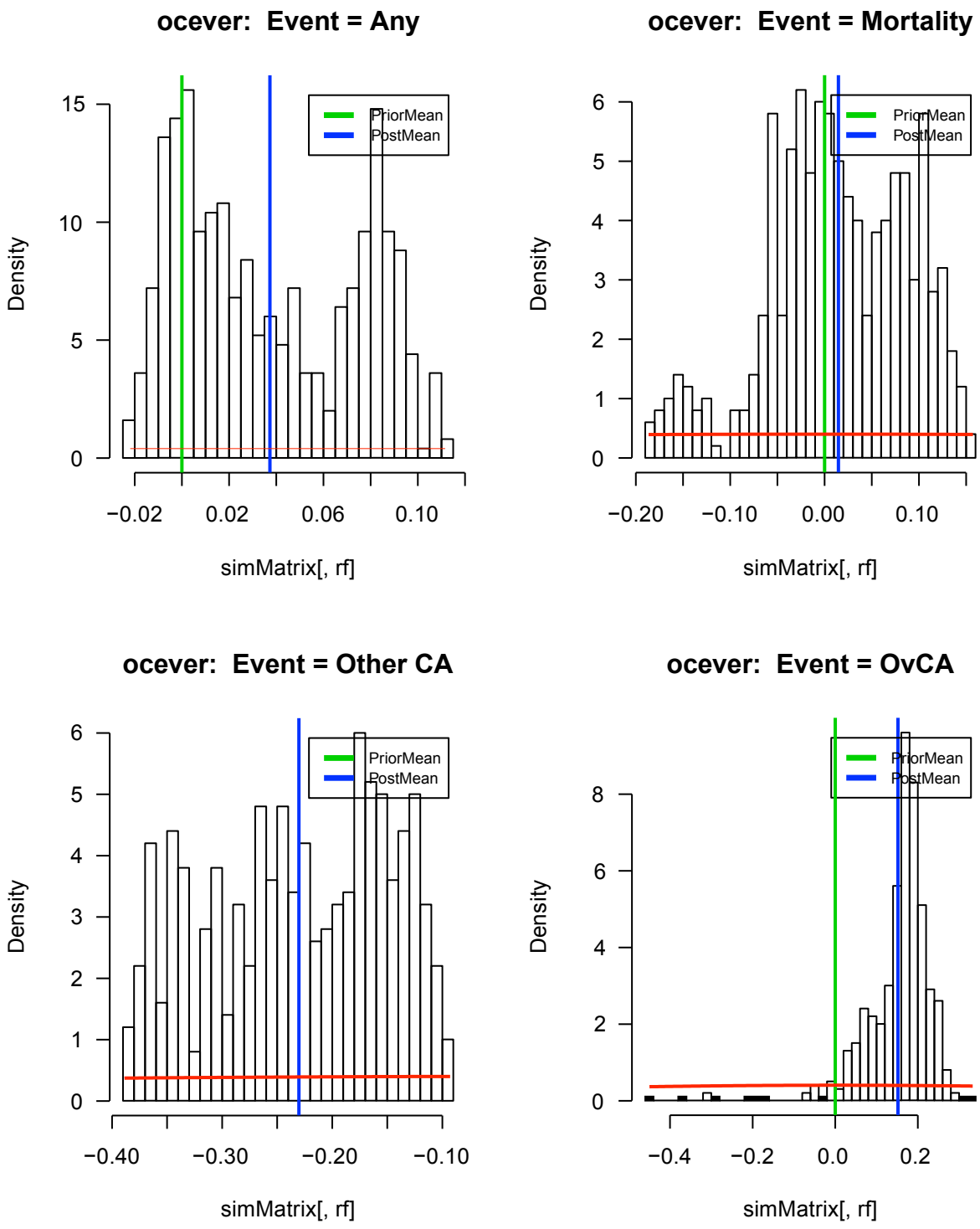

delta0[7]: Study = NJO
PriorMean
PostMean
Density
simMatrix[, pnames[i]]

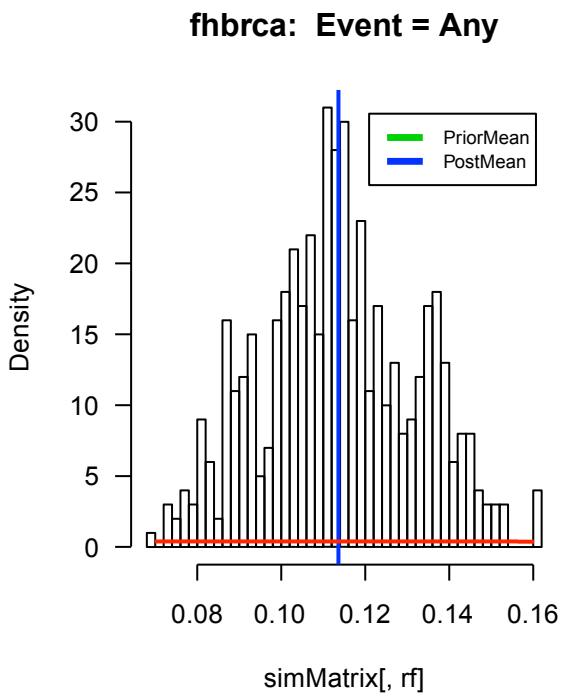

delta0[8]: Study = STA
PriorMean
PostMean
Density
simMatrix[, pnames[i]]

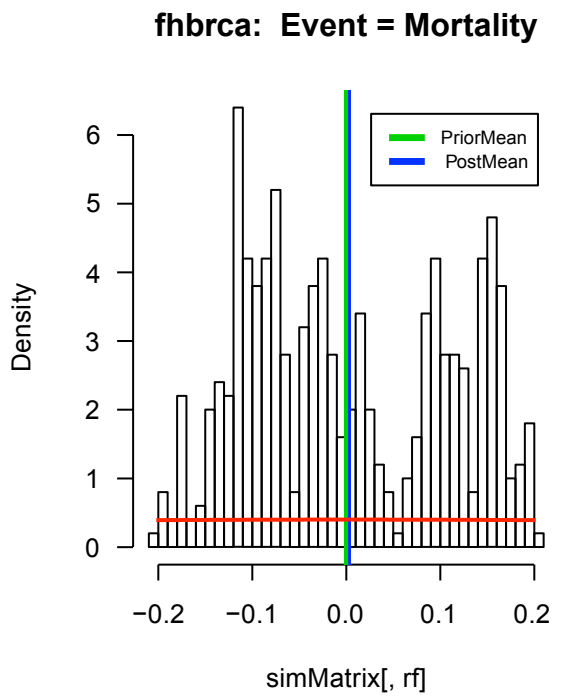

delta0[9]: Study = UCI
PriorMean
PostMean
Density
simMatrix[, pnames[i]]

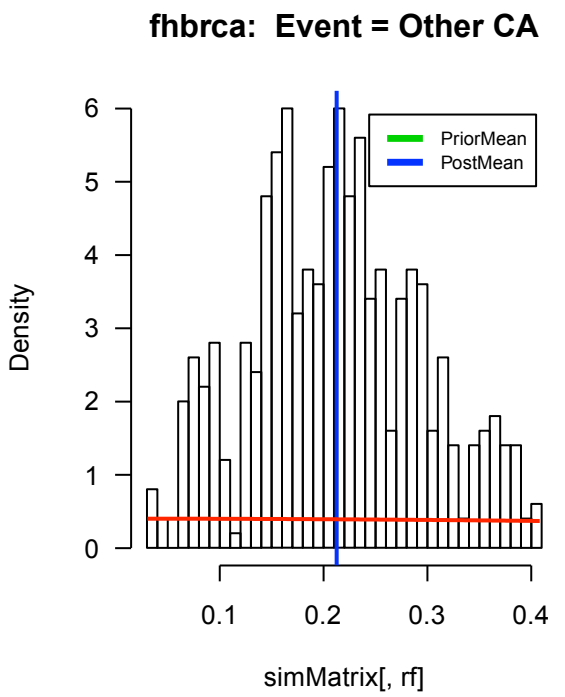

delta0[10]: Study = USC
PriorMean
PostMean
Density
simMatrix[, pnames[i]]

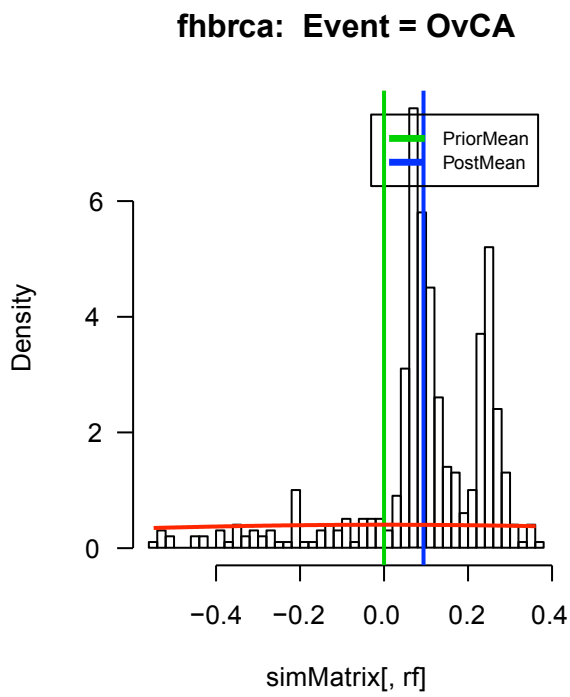

sd.ovca: Study = NA
PriorMean
PostMean
Density
simMatrix[, pnames[i]]

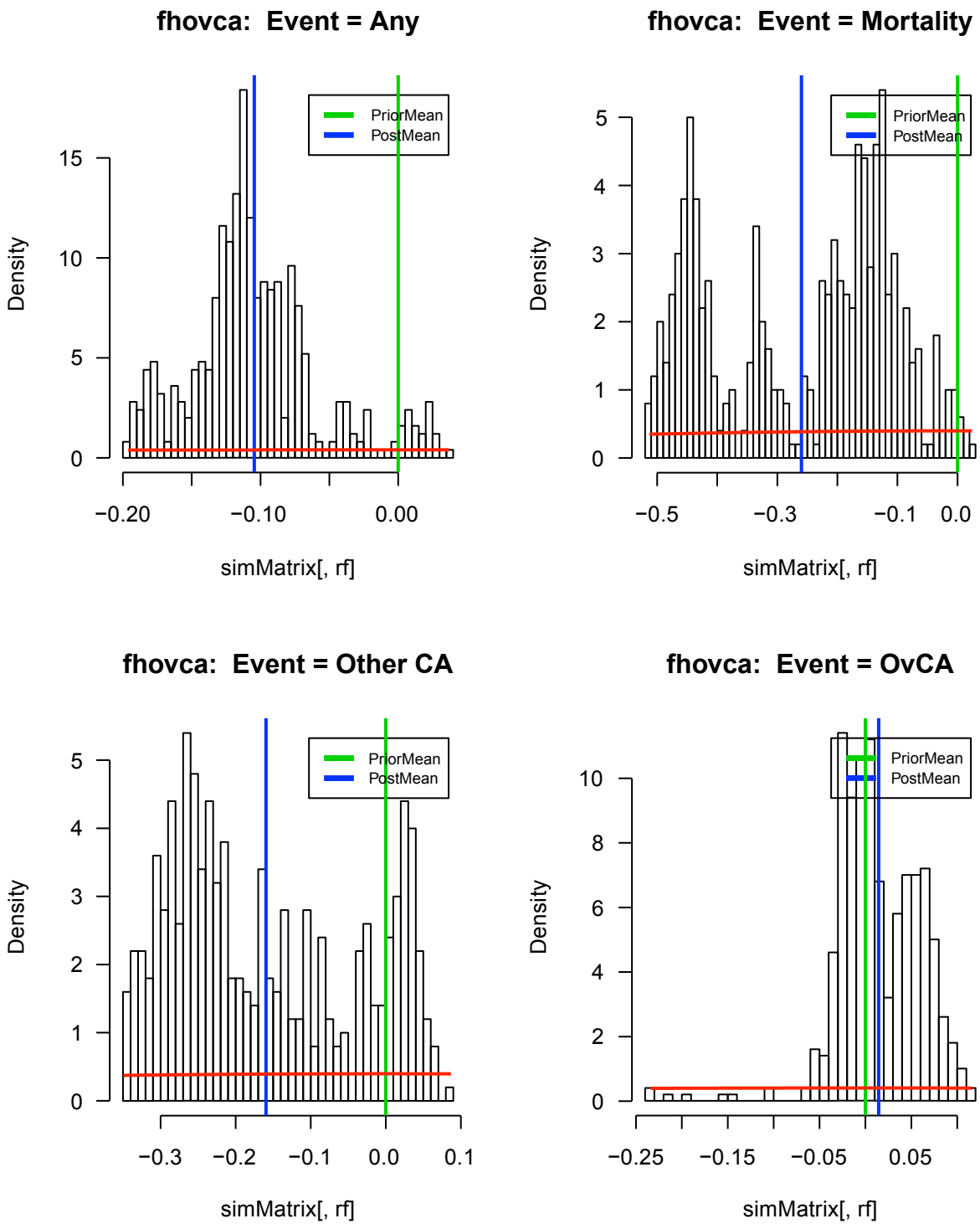

ocever: Event = Any
PriorMean
PostMean
Density
simMatrix[, rf]
ocever: Event = Mortality
PriorMean
PostMean
Density
simMatrix[, rf]
ocever: Event = Other CA
PriorMean
PostMean
Density
simMatrix[, rf]
ocever: Event = OvCA
PriorMean
PostMean
Density
simMatrix[, rf]

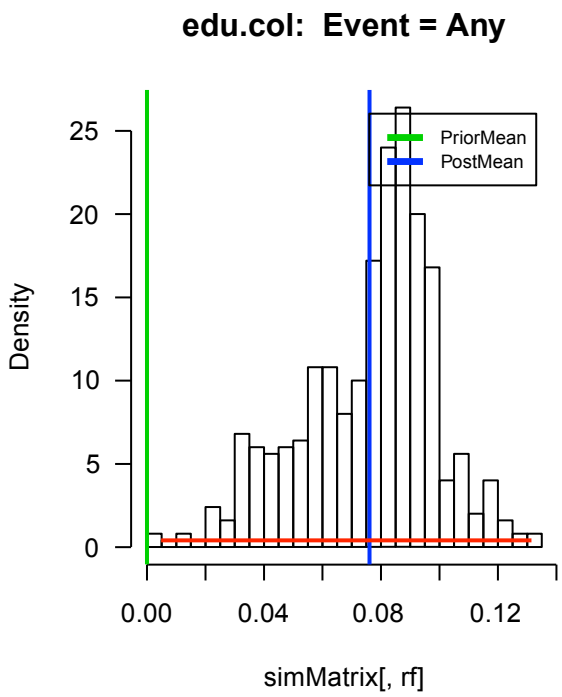

fhbrca: Event = Any
PriorMean
PostMean
Density
simMatrix[, rf]

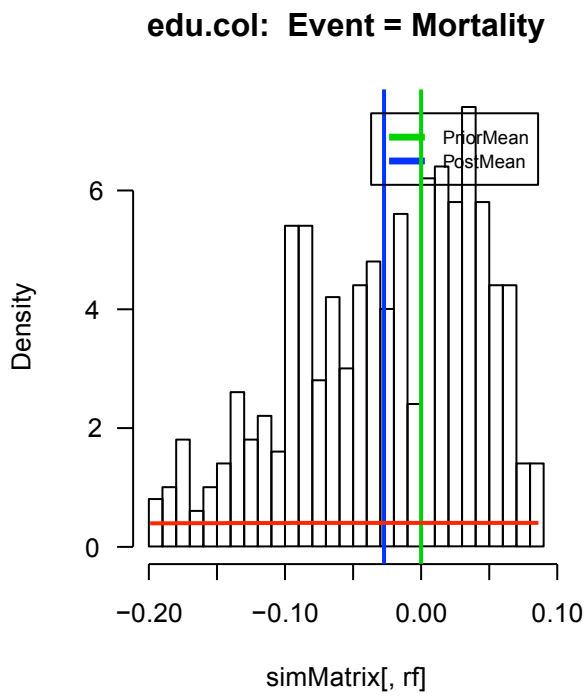

fhbrca: Event = Mortality
PriorMean
PostMean
Density
simMatrix[, rf]

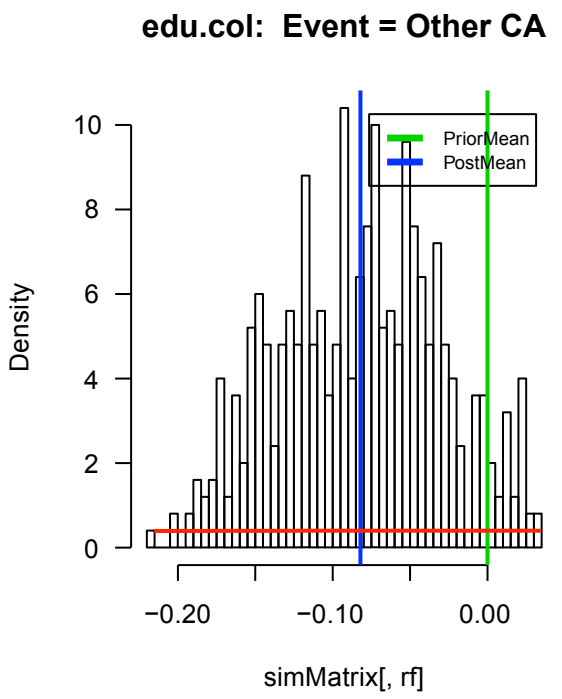

fhbrca: Event = Other CA
PriorMean
PostMean
Density
simMatrix[, rf]

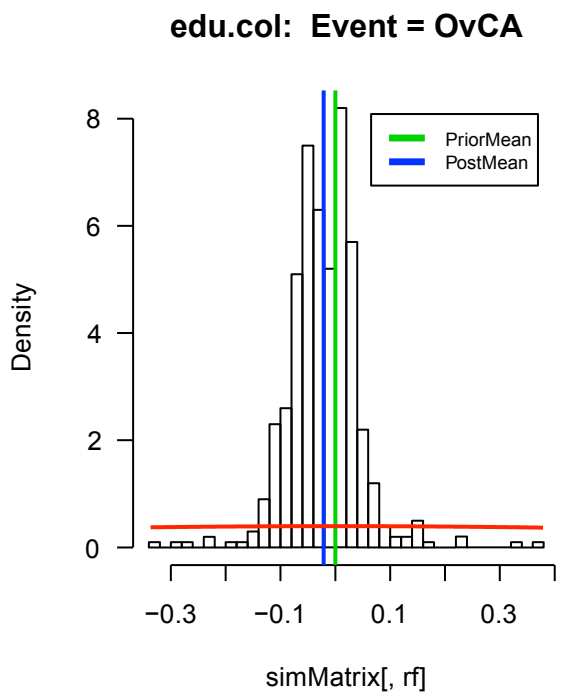

fhbrca: Event = OvCA
PriorMean
PostMean
Density
simMatrix[, rf]

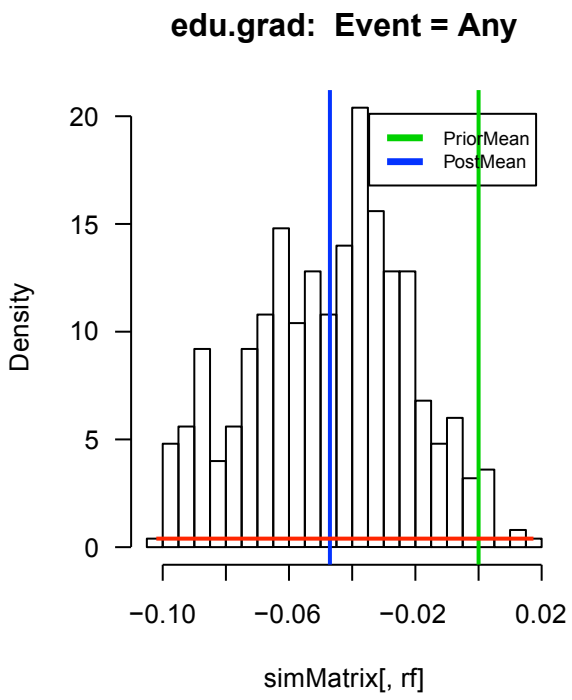

fhovca: Event = Any
PriorMean
PostMean
Density
simMatrix[, rf]
fhovca: Event = Mortality
PriorMean
PostMean
Density
simMatrix[, rf]
fhovca: Event = Other CA
PriorMean
PostMean
Density
simMatrix[, rf]
fhovca: Event = OvCA
PriorMean
PostMean
Density
simMatrix[, rf]

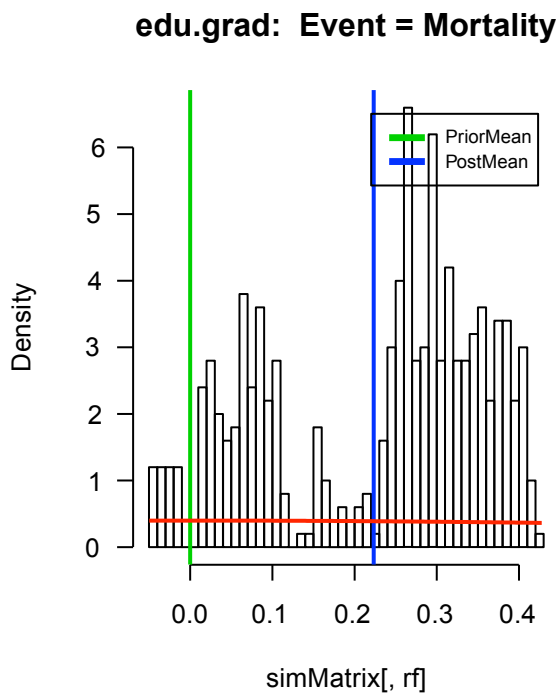

edu.col: Event = Any
PriorMean
PostMean
Density
simMatrix[, rf]

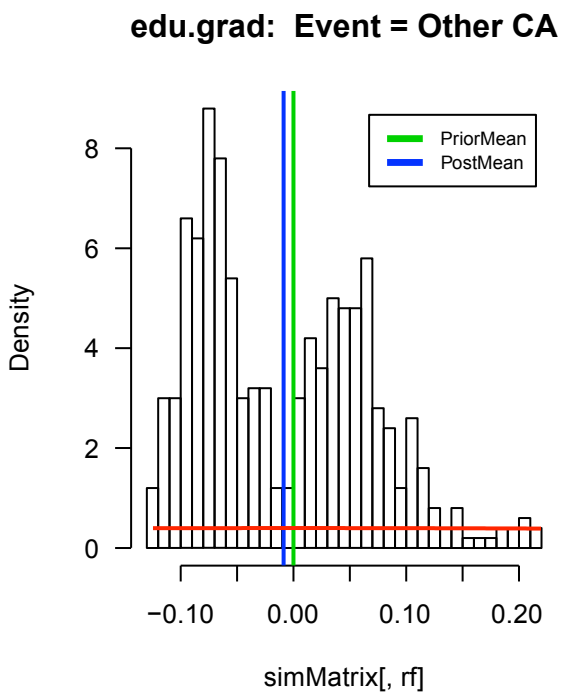

edu.col: Event = Mortality
PriorMean
PostMean
Density
simMatrix[, rf]

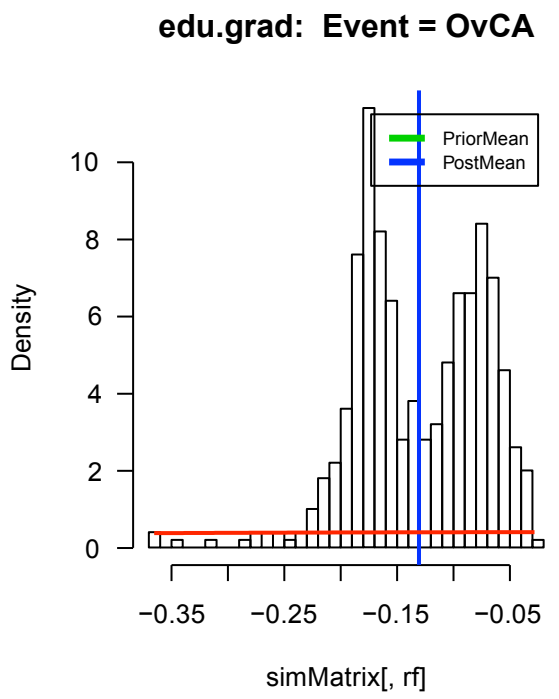

edu.col: Event = Other CA
PriorMean
PostMean
Density
simMatrix[, rf]

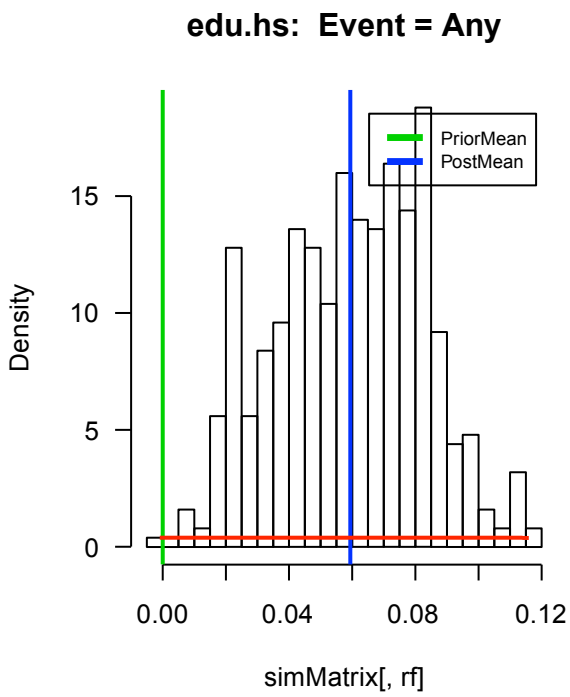

edu.col: Event = OvCA
PriorMean
PostMean
Density
simMatrix[, rf]

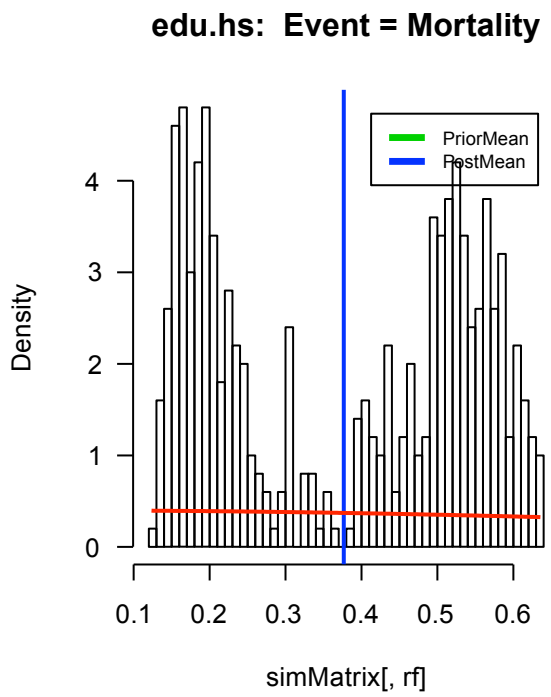

edu.grad: Event = Any
PriorMean
PostMean
Density
simMatrix[, rf]

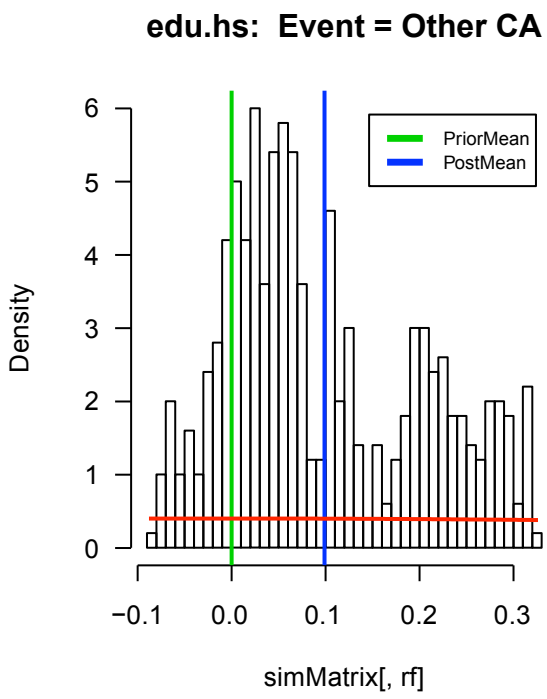

edu.grad: Event = Mortality
PriorMean
PostMean
Density
simMatrix[, rf]

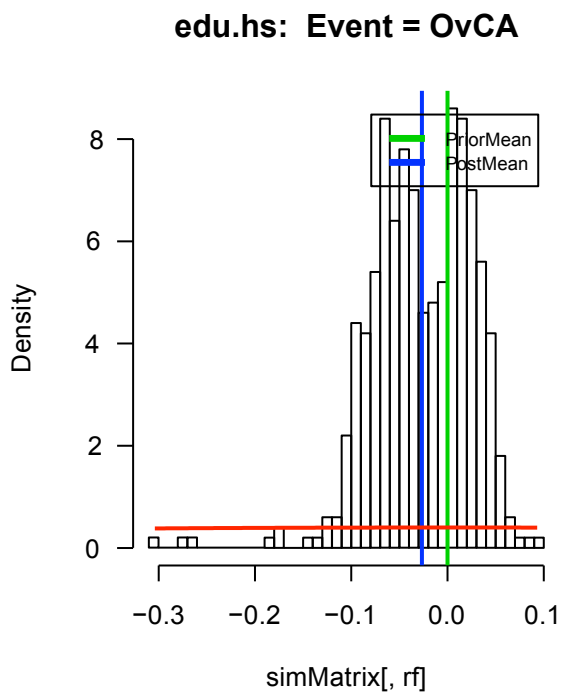

edu.grad: Event = Other CA
PriorMean
PostMean
Density
simMatrix[, rf]

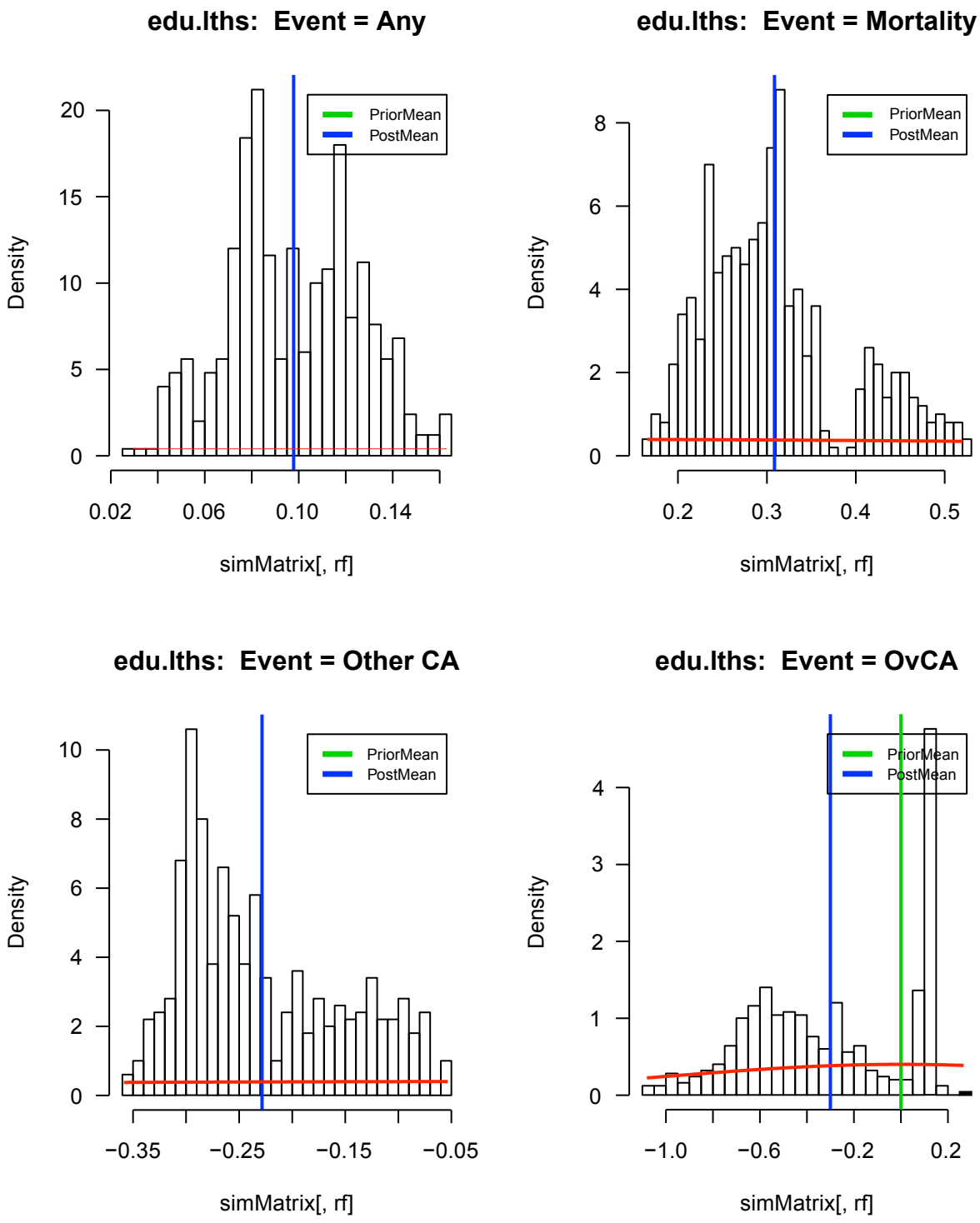

edu.grad: Event = OvCA
PriorMean
PostMean
Density
simMatrix[, rf]

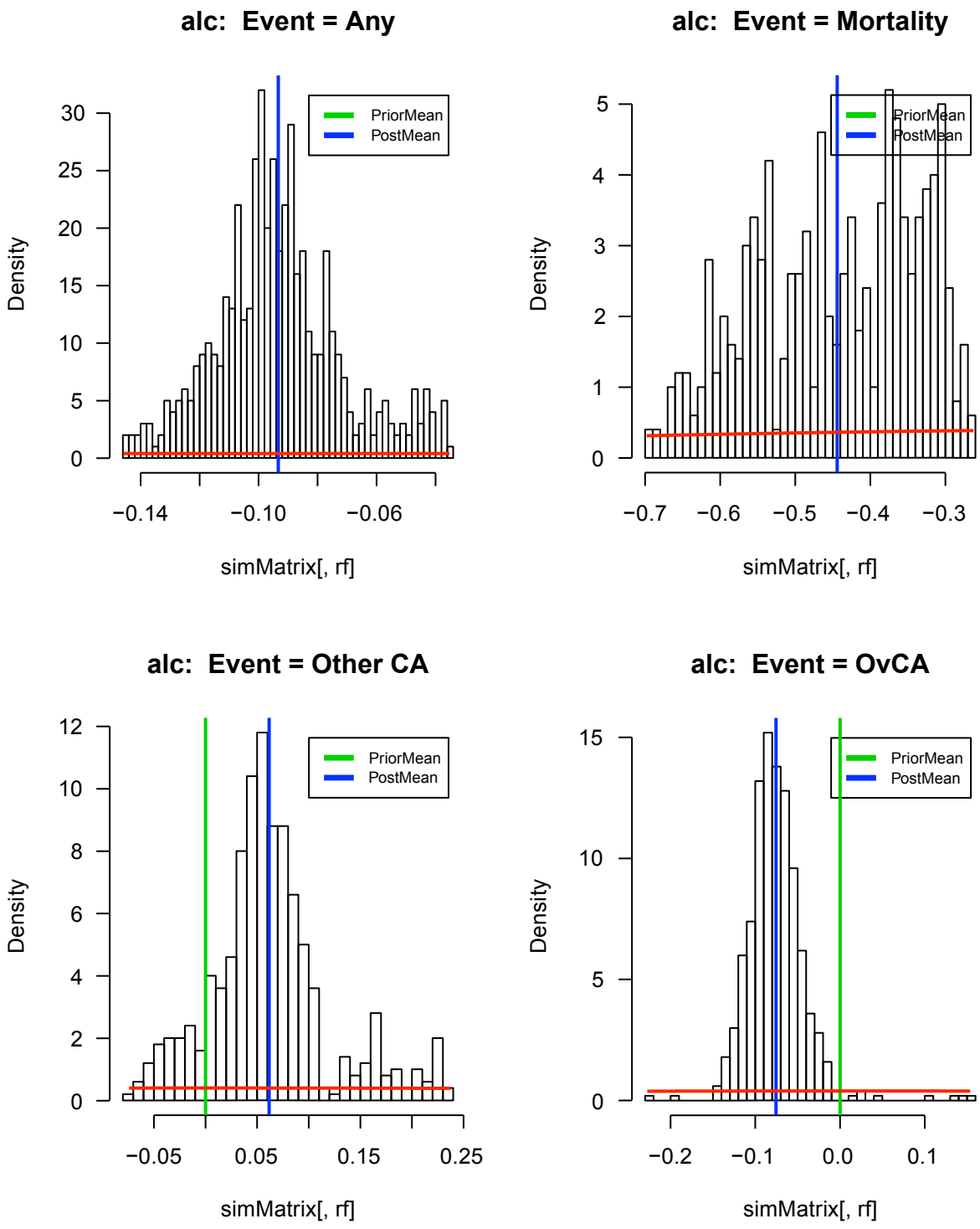

edu.hs: Event = Any
PriorMean
PostMean
Density
simMatrix[, rf]

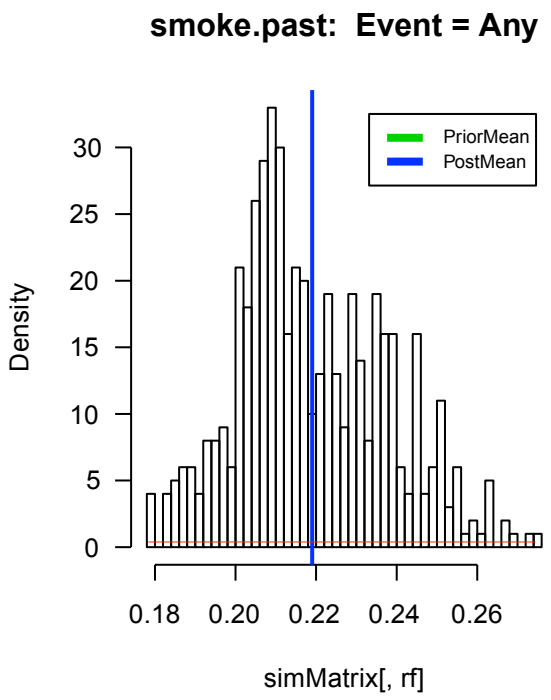

edu.hs: Event = Mortality
PriorMean
PostMean
Density
simMatrix[, rf]

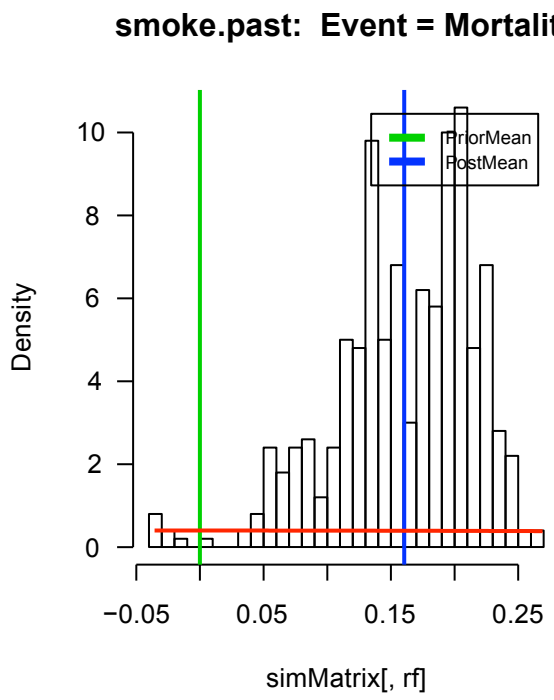

edu.hs: Event = Other CA
PriorMean
PostMean
Density
simMatrix[, rf]

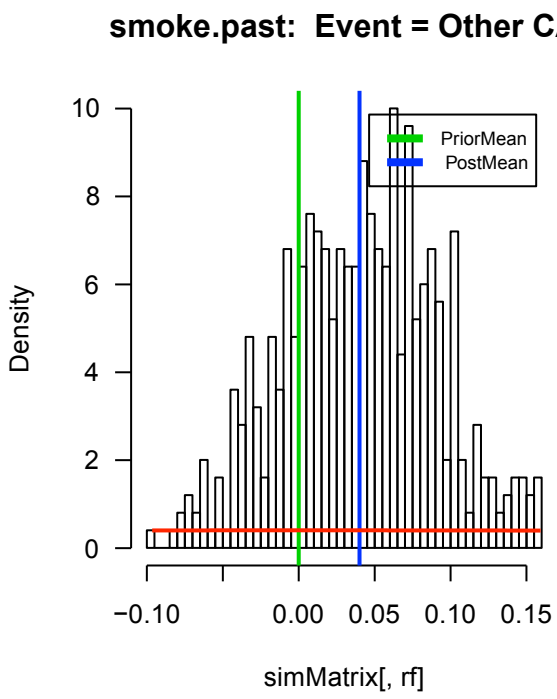

edu.hs: Event = OvCA
PriorMean
PostMean
Density
simMatrix[, rf]

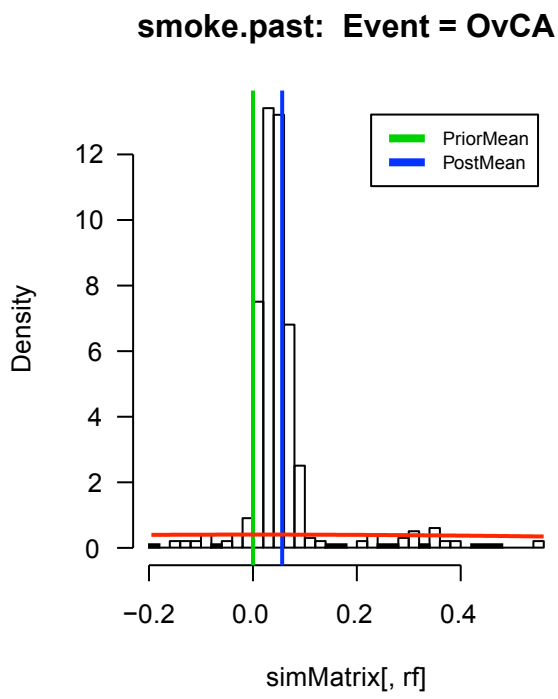

edu.lths: Event = Any
PriorMean
PostMean
Density
simMatrix[, rf]
edu.lths: Event = Mortality
PriorMean
PostMean
Density
simMatrix[, rf]
edu.lths: Event = Other CA
PriorMean
PostMean
Density
simMatrix[, rf]
edu.lths: Event = OvCA
PriorMean
PostMean
Density
simMatrix[, rf]

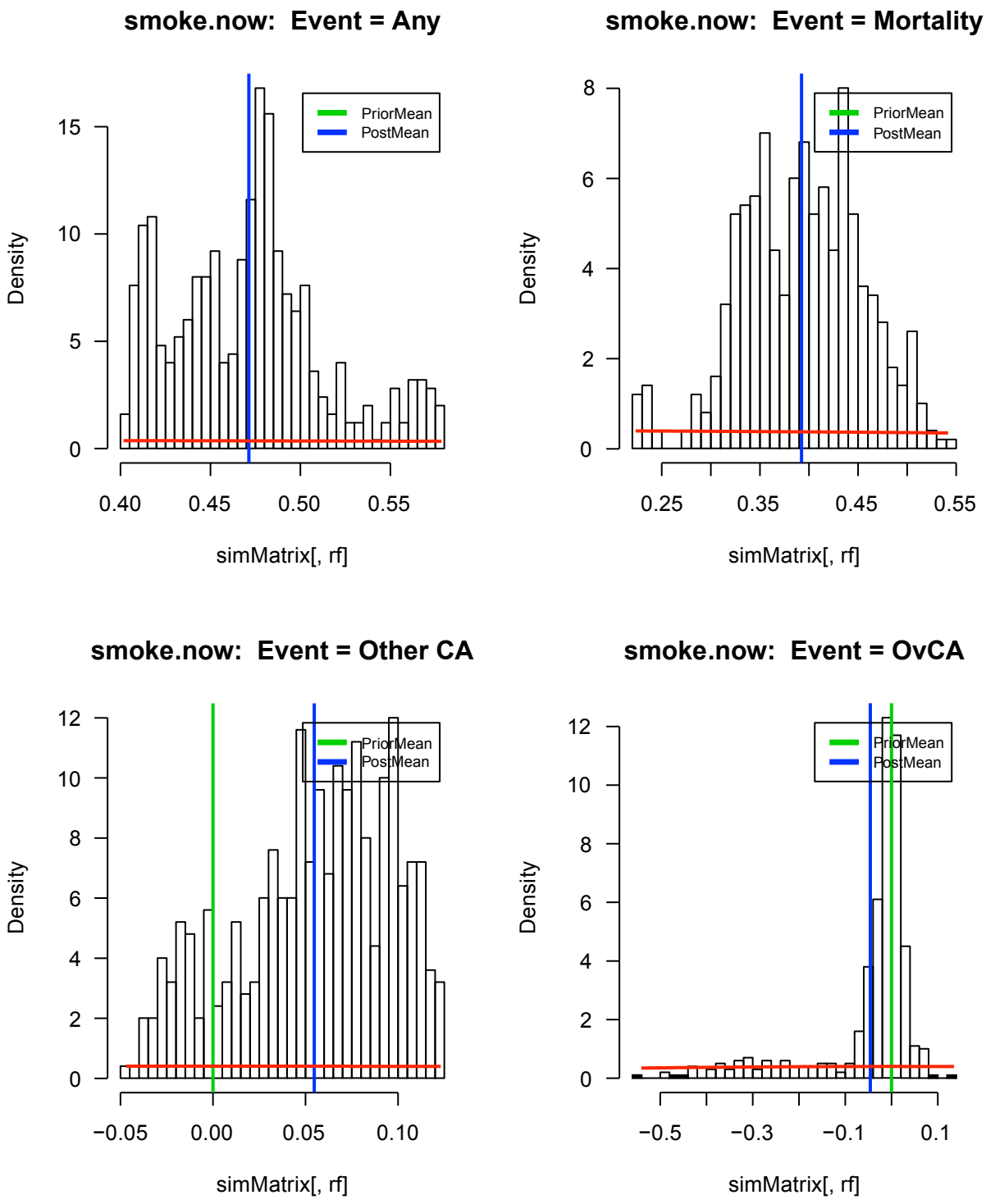

alc: Event = Any
PriorMean
PostMean
Density
simMatrix[, rf]
alc: Event = Mortality
PriorMean
PostMean
Density
simMatrix[, rf]
alc: Event = Other CA
PriorMean
PostMean
Density
simMatrix[, rf]
alc: Event = OvCA
PriorMean
PostMean
Density
simMatrix[, rf]

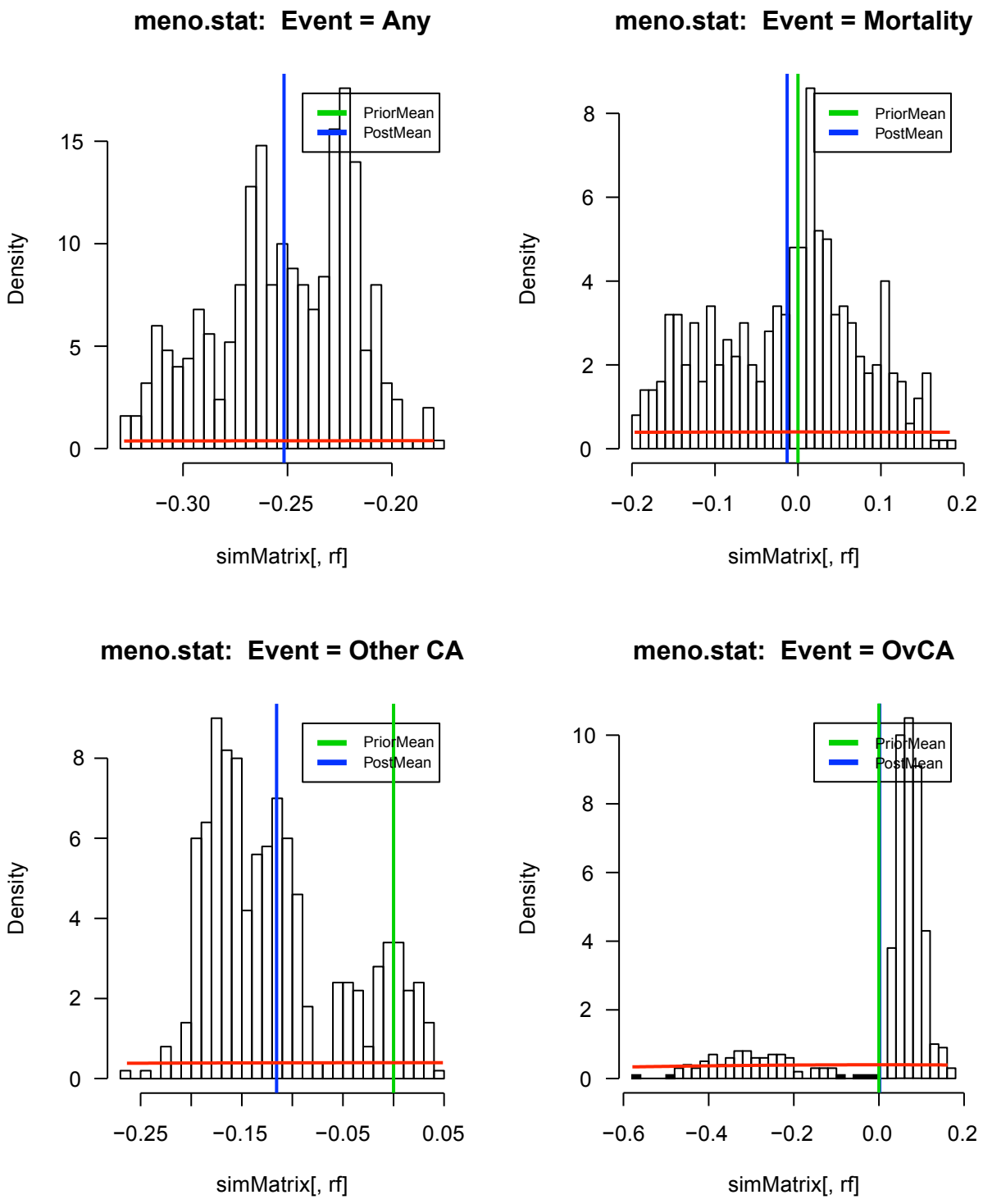

smoke.past: Event = Any
PriorMean
PostMean
Density
simMatrix[, rf]

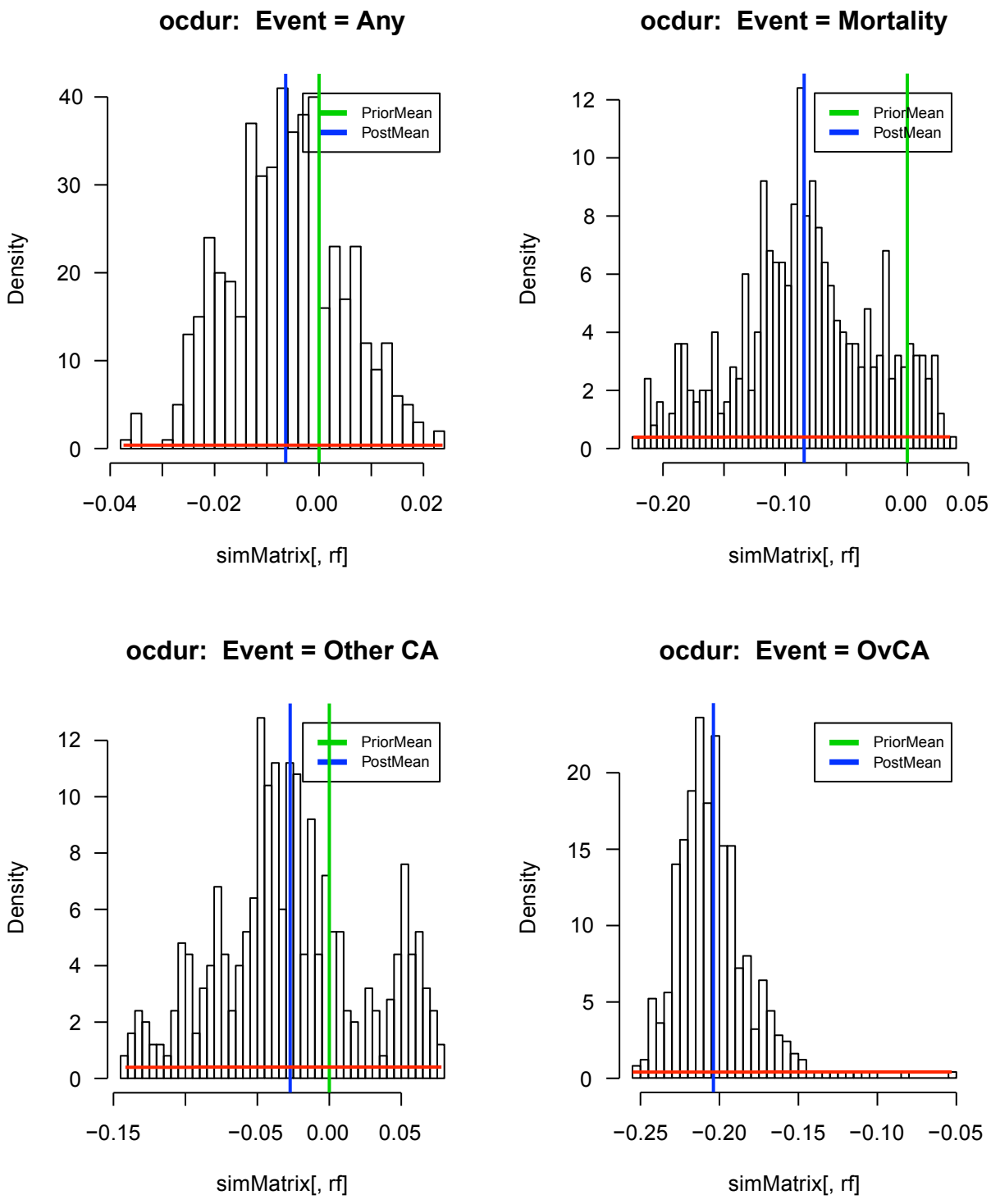

smoke.past: Event = Mortality
PriorMean
PostMean
Density
simMatrix[, rf]

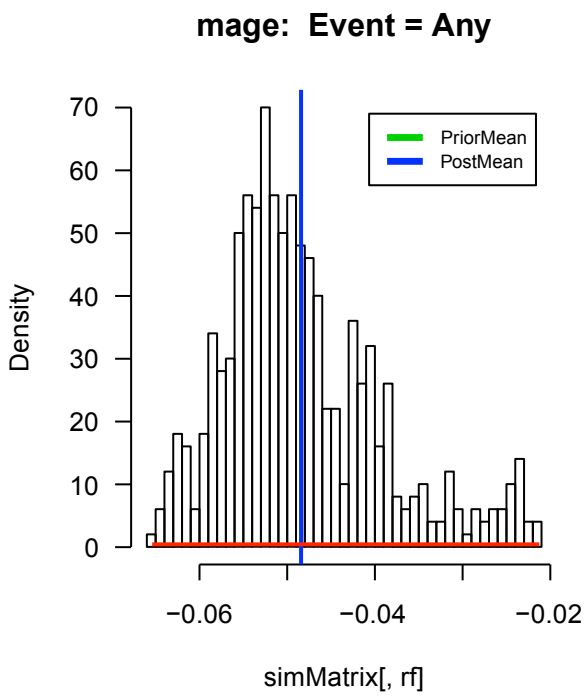

smoke.past: Event = Other CA
PriorMean
PostMean
Density
simMatrix[, rf]

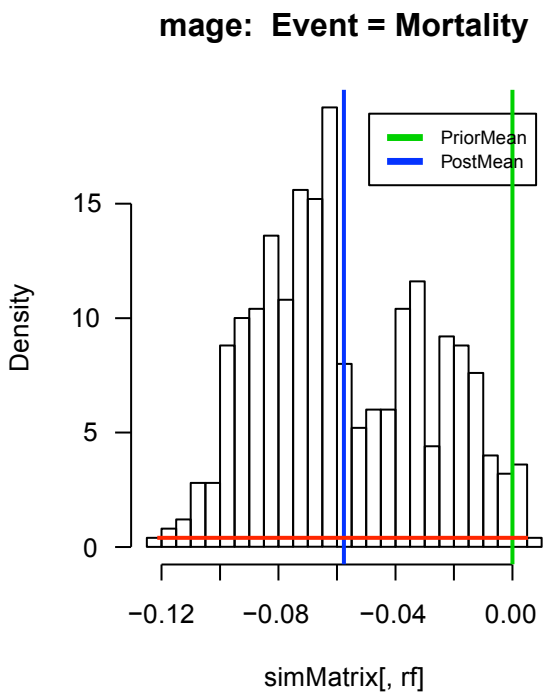

smoke.past: Event = OvCA
PriorMean
PostMean
Density
simMatrix[, rf]

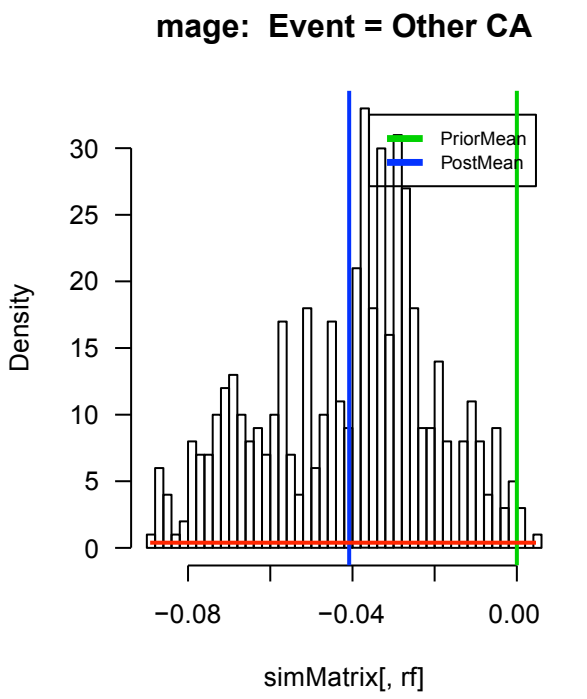

smoke.now: Event = Any
PriorMean
PostMean
Density
simMatrix[, rf]
smoke.now: Event = Mortality
PriorMean
PostMean
Density
simMatrix[, rf]
smoke.now: Event = Other CA
PriorMean
PostMean
Density
simMatrix[, rf]
smoke.now: Event = OvCA
PriorMean
PostMean
Density
simMatrix[, rf]

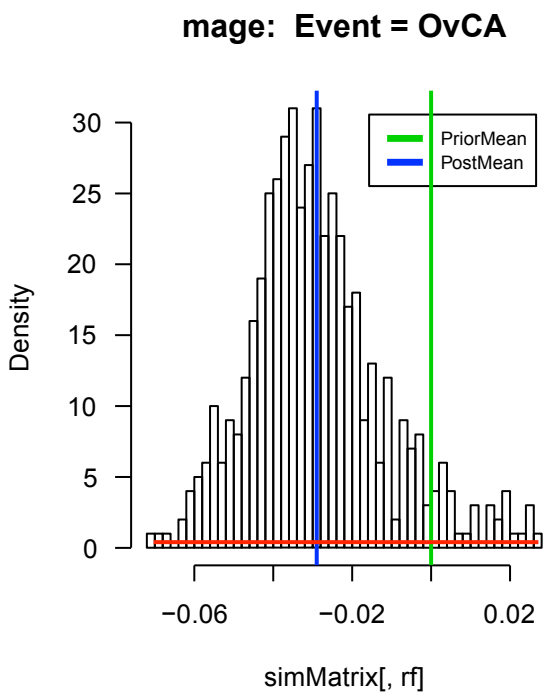

meno.stat: Event = Any
PriorMean
PostMean
Density
simMatrix[, rf]
meno.stat: Event = Mortality
PriorMean
PostMean
Density
simMatrix[, rf]
meno.stat: Event = Other CA
PriorMean
PostMean
Density
simMatrix[, rf]
meno.stat: Event = OvCA
PriorMean
PostMean
Density
simMatrix[, rf]

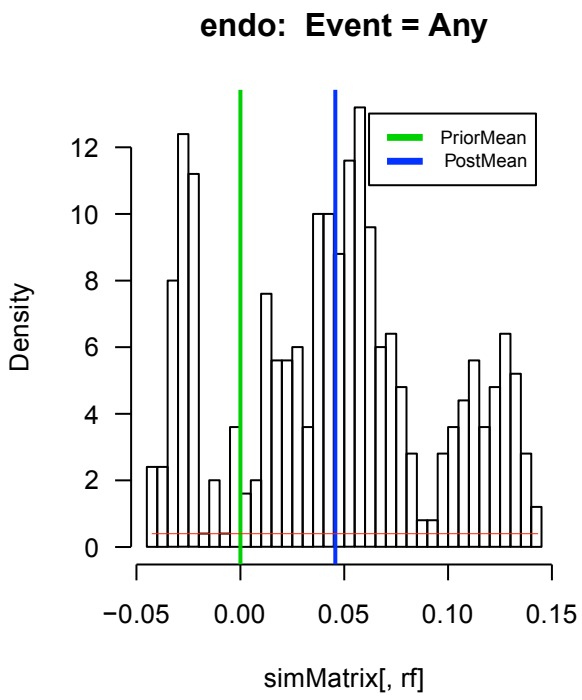

ocdur: Event = Any
ocdur: Event = Mortality
ocdur: Event = Other CA
ocdur: Event = OvCA
PriorMean
PostMean
Density
simMatrix[, rf]

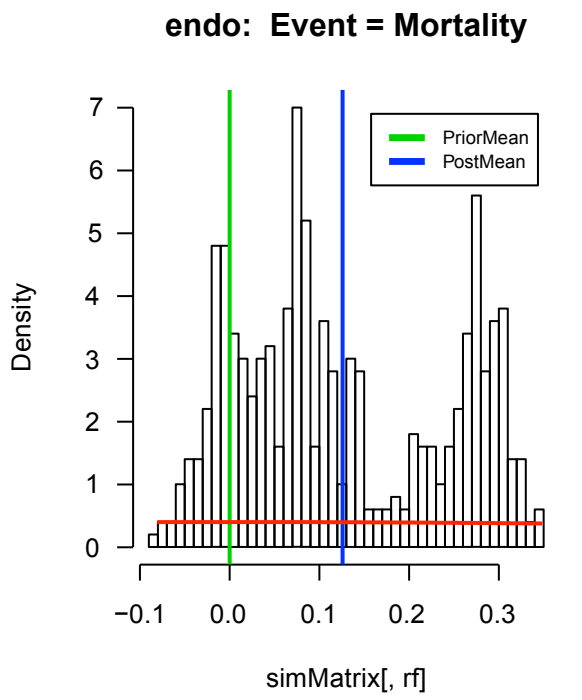

mage: Event = Any
PriorMean
PostMean
Density
simMatrix[, rf]

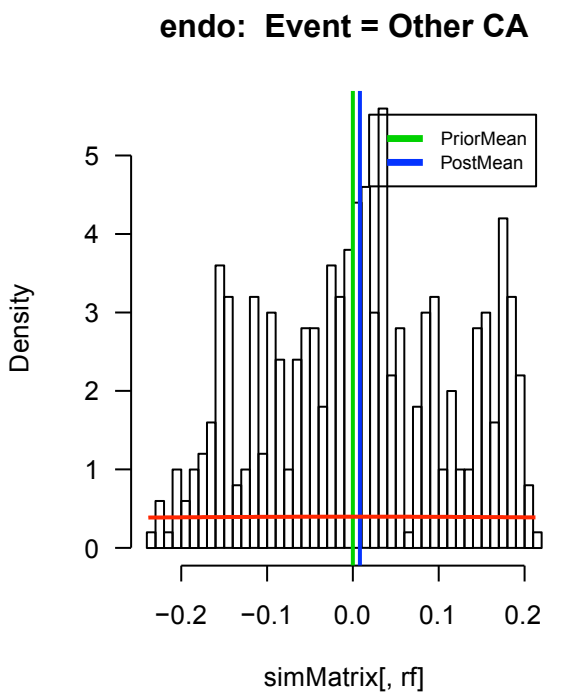

mage: Event = Mortality
PriorMean
PostMean
Density
simMatrix[, rf]

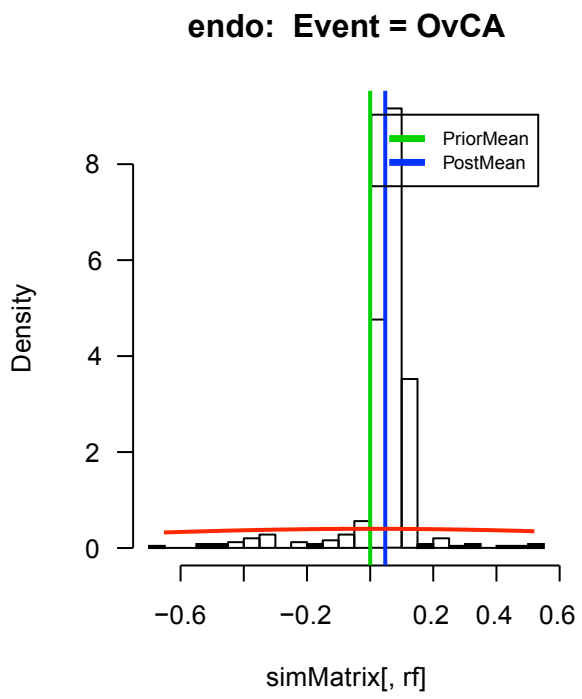

mage: Event = Other CA
PriorMean
PostMean
Density
simMatrix[, rf]

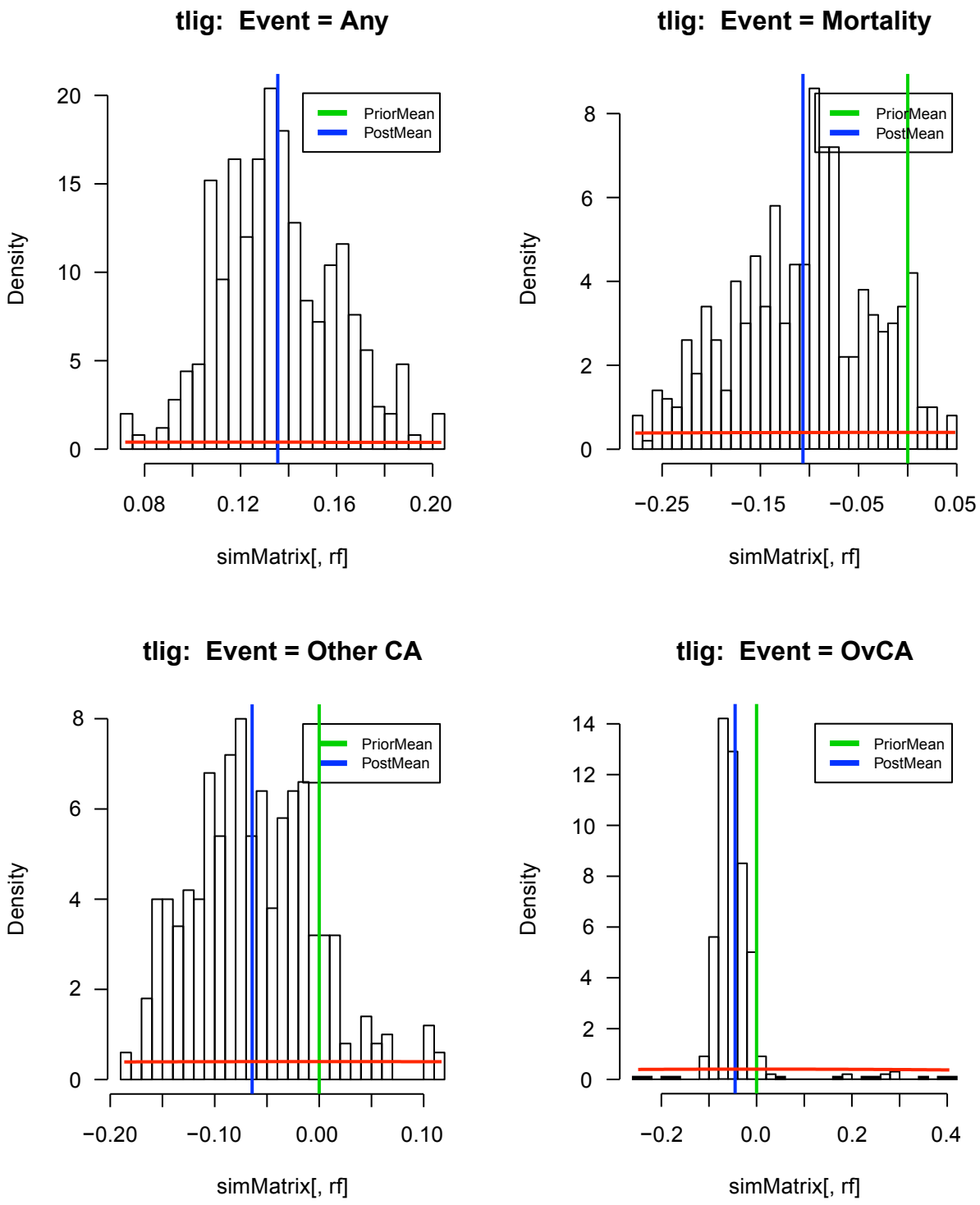

mage: Event = OvCA
PriorMean
PostMean
Density
simMatrix[, rf]


endo: Event = Any
PriorMean
PostMean
Density
simMatrix[, rf]


endo: Event = Mortality
PriorMean
PostMean
Density
simMatrix[, rf]


endo: Event = Other CA
PriorMean
PostMean
Density
simMatrix[, rf]


endo: Event = OvCA
PriorMean
PostMean
Density
simMatrix[, rf]




```
rm(rfplots)
## Trace Plots:
```

```
pdf("TracePlot.pdf", height = 9, width = 6.5)
par(mfrow = c(4, 1), las = 1) cnames <-
colnames(simMatrix) for (i in
1:ncol(simMatrix)) {
    plot(simMatrix[, i], las = 1, main = paste("Trace of ", cnames[i],
        sep = ""))
}
dev.off()
```

```
## pdf
##   2
```

2.0.1 Wrap–Up

```
gc()
```

```
##           used (Mb) gc trigger (Mb) max used (Mb)
## Ncells  472477 25.3    940480 50.3   940480 50.3
## Vcells 27347683 208.7  59615938 454.9 59613884 454.9 save.image()
```

Appendix 4: Results for androgen and IGF-1 concentrations and risk of ovarian cancer by histology

Odds ratios (95% CI) for invasive EOC for doubling and by tertiles in the OC3 [1]

		Invasive	
	Sets	**OR 95% CI**	**p_{trend}**
Testosterone			
Q1	398	ref	
Q2	443	**1.18 (1.00 - 1.39)**	
Q3[2]	460	**1.24 (1.04 - 1.47)**	**0.03**
Doubling[3]	1,301	**1.12 (1.01 - 1.23)**	**0.03**
Free Testosterone			
Q1	286	ref	
Q2	287	1.03 (0.84 - 1.26)	
Q3[2]	292	1.05 (0.86 - 1.29)	0.05
Doubling[3]	865	1.10 (1.00 - 1.21)	0.06
Androstenedione			
Q1	450	ref	
Q2	387	0.86 (0.72 - 1.02)	
Q3[2]	470	1.07 (0.89 - 1.27)	0.15
Doubling[3]	1,307	1.07 (0.97 - 1.19)	0.19
DHEAS			
Q1	213	ref	
Q2	225	1.05 (0.83 - 1.33)	
Q3[2]	207	0.95 (0.74 - 1.23)	0.92
Doubling[3]	645	0.98 (0.88 - 1.10)	0.76
SHBG			
Q1	311	ref	
Q2	250	0.82 (0.67 - 1.00)	
Q3[2]	325	1.08 (0.88 - 1.31)	0.85
Doubling[3]	886	1.02 (0.91 - 1.14)	0.77

IGF			
Q1	460	ref	
Q2	424	0.91 (0.78 - 1.08)	
Q3[2]	386	**0.81 (0.68 - 0.97)**	**<0.01**
Doubling[3]	1,270	**0.82 (0.73 - 0.93)**	**<0.01**

[1]Results were derived from conditional logistic regression models, additionally adjusted for OC use (never/ever/missing) and parity (never/ever/missing).
[2]The p value for trend across tertiles is based on a continuous probit score (generating a rank for each person in each cohort by hormone level). [3]Linear trends for doubling of hormone concentrations estimated on $\log_2$ scale.
[4] Pair-wise heterogeneity tests were performed, using the likelihood ratio test comparing models assuming (1) the same association between exposure and outcomes compared to (2) a model assuming different associations for each subtype.

Odds ratios (95% CI) for EOC by histological subtypes for doubling and by tertiles in the OC3[1]

	Serous			Endometrioid			Mucinous			Clear Cell			
	Sets	**OR 95%CI**	**p_{trend}**	**Sets**	**OR 95%CI**	**p_{trend}**	**Sets**	**OR 95%CI**	**p_{trend}**	**Sets**	**OR 95%CI**	**p_{trend}**	**p_{het}**
Testosterone													
Q1	222	ref		35	ref		45	ref		15	ref		
Q2	229	1.16 (0.92 - 1.46)		60	1.44 (0.87 - 2.38)		61	1.34 (0.86 - 2.08)		27	1.50 (0.67 - 3.37)		
Q3[2]	204	0.96 (0.76 - 1.23)	0.51	69	**1.78 (1.07 - 2.98)**	0.07	84	**1.95 (1.25 - 3.03)**	0.05	17	0.73 (0.30 - 1.77)	0.82	
Doubling[3]	655	0.96 (0.83 - 1.10)	0.57	164	**1.39 (1.02 - 1.89)**	**0.04**	190	**1.29 (1.01 - 1.66)**	**0.04**	59	1.06 (0.67 - 1.69)	0.80	**0.02**
Free Testosterone													
Q1	155	ref		25	ref		35	ref		11	ref		
Q2	151	1.03 (0.78 - 1.36)		32	0.99 (0.51 - 1.93)		48	1.46 (0.85 - 2.52)		10	0.85 (0.29 - 2.48)		
Q3[2]	129	0.82 (0.62 - 1.10)	0.64	36	1.00 (0.52 - 1.93)	0.54	50	1.49 (0.88 - 2.52)	**0.03**	20	1.98 (0.68 - 5.78)	0.28	
Doubling[3]	435	0.97 (0.84 - 1.11)	0.64	93	1.09 (0.79 - 1.5)	0.62	133	**1.33 (1.03 - 1.71)**	**0.03**	41	1.28 (0.81 - 2.02)	0.28	0.06
Androstenedione													
Q1	235	ref		46	ref		56	ref		21	ref		
Q2	204	0.88 (0.70 - 1.11)		45	0.70 (0.42 - 1.18)		51	0.90 (0.57 - 1.42)		15	0.80 (0.36 - 1.77)		
Q3[2]	217	0.99 (0.77 - 1.27)	0.72	73	1.02 (0.61 - 1.71)	0.80	84	**1.57 (1.02 - 2.43)**	**0.03**	24	0.75 (0.36 - 1.60)	0.75	
Doubling[3]	656	0.96 (0.83 - 1.11)	0.62	164	1.04 (0.76 - 1.43)	0.79	191	**1.33 (1.03 - 1.72)**	**0.03**	60	1.02 (0.67 - 1.55)	0.94	0.26
DHEAS													
Q1	123	ref		17	ref		8	ref		7	ref		
Q2	120	0.98 (0.71 - 1.35)		20	0.71 (0.30 - 1.65)		14	1.04 (0.35 - 3.1)		11	2.31 (0.65 - 8.21)		
Q3[2]	104	0.78 (0.55 - 1.11)	0.31	27	1.06 (0.43 - 2.61)	0.82	20	1.71 (0.55 - 5.33)	0.21	13	2.49 (0.74 - 8.37)	0.31	
Doubling[3]	347	0.92 (0.79 - 1.06)	0.24	64	1.06 (0.72 - 1.55)	0.78	42	1.42 (0.84 - 2.42)	0.19	31	1.24 (0.71 - 2.17)	0.45	0.56
SHBG													

Q1	147	ref		30	ref		56	ref		19	ref		
Q2	141	0.87 (0.66 - 1.16)		27	1.07 (0.55 - 2.08)		28	0.61 (0.36 - 1.04)		12	0.72 (0.29 - 1.79)		
Q3[2]	157	1.12 (0.85 - 1.48)	0.71	37	1.51 (0.80 - 2.87)	0.63	51	0.95 (0.58 - 1.55)	0.78	12	0.66 (0.26 - 1.67)	0.30	
Doubling[3]	445	1.06 (0.91 - 1.24)	0.46	94	1.17 (0.81 - 1.68)	0.40	135	0.93 (0.68 - 1.28)	0.66	43	0.74 (0.45 - 1.22)	0.23	0.24
IGF													
Q1	211	ref		56	ref		67	ref		25	ref		
Q2	209	0.95 (0.75 - 1.21)		48	0.77 (0.46 - 1.28)		64	0.91 (0.60 - 1.39)		20	0.70 (0.33 - 1.52)		
Q3[2]	209	0.94 (0.73 - 1.21)	0.21	59	0.81 (0.48 - 1.38)	0.34	55	0.75 (0.48 - 1.19)	0.21	12	0.50 (0.21 - 1.23)	0.07	
Doubling[3]	630	0.89 (0.74 - 1.06)	0.20	163	0.83 (0.57 - 1.22)	0.35	186	0.81 (0.58 - 1.13)	0.21	57	0.55 (0.29 - 1.05)	0.07	0.71

[1]Results were derived from conditional logistic regression models, additionally adjusted for OC use (never/ever) and parity (never/ever).
[2]The p value for trend across tertiles is based on a continuous probit score (generating a rank for each person in each cohort by hormone level).
[3]Linear trends for doubling of hormone concentrations estimated on $\log_2$ scale.
[4] Pair-wise heterogeneity tests were performed, using the likelihood ratio test comparing models assuming (1) the same association between exposure and outcomes compared to (2) a model assuming different associations for each subtype.

Appendix 5: Submitted aims for an R01 using the OC3 to examine inflammation and ovarian cancer risk

Ovarian cancer is the fifth leading cause of cancer death in the US.[1,2] Few ovarian cancer risk factors (e.g., pregnancy) are easily modifiable,[3] thus it is critical to identify new, potentially modifiable/treatable risk factors to improve prevention. Further, established risk factors show different associations by tumor subtypes,[4-7] with few being associated with aggressive disease (e.g., serous, death within 3 years). This highlights two critical needs in ovarian cancer research: (1) consortia to accrue enough well-characterized cases to assess associations by tumor subtypes and (2) identification of pathways that drive the development of aggressive tumors. Here, we propose to comprehensively characterize the role of inflammation, a modifiable exposure, in ovarian cancer leveraging, and expanding the resources of, the Ovarian Cancer Cohort Consortium (OC3), a collaboration of 23 cohorts with >8,000 ovarian cancer cases (~1500 with biomarker data) in 1.5 million women

Increasing evidence supports inflammation as a key mechanism in ovarian cancer; however, questions remain. Ovarian tumors are characterized by dysregulation of interleukin (IL)-6 and tumor necrosis factor (TNF) α;[8-12] patients with high circulating IL-6 and TNFα have worse survival, suggesting inflammation may be related to aggressive disease.[13,14] However, prospective studies evaluating circulating levels of these markers have been mixed, although most were small.[15-18] Conversely, despite a lack of biologic data supporting C-reactive protein (CRP) in ovarian tumorigenesis, pre-diagnosis CRP has been consistently positively associated with ovarian cancer risk, particularly for overweight women.[16-21] However, CRP is non-specific, and as it likely reflects other inflammatory processes that promote carcinogenesis, it may not directly impact ovarian cancer risk. That said, factors that increase CRP (e.g., smoking[22]) are not strongly related to overall ovarian cancer risk.[23] Further, CRP, IL-6, and TNFα are increased by ovarian tumors,[24,25] leading to the potential for reverse causation. Thus, ovarian cancer research would be greatly enhanced by assessing: (1) novel inflammatory exposures, (2) combining biomarkers or exposures to reflect overall inflammatory profiles, since each likely explains only a small portion of the variation in inflammation relevant for ovarian cancer, and (3) if associations are stronger for aggressive disease and persist over follow-up. Thus, we propose to evaluate circulating CRP, IL-6, and TNFα-R2 (a marker of TNFα activation), their genetic predictors, and a wide range of inflammatory exposures with ovarian cancer risk overall and by tumor subtype, including immunohistochemical (IHC) subtyping, and to consider if grouping exposures that define inflammatory profiles highlights pathways for prevention.

Currently the OC3 includes baseline exposure data and disease follow-up for up to 35 years. To implement this proposal, we will incorporate biomarker data from serum/plasma, DNA, and, in a pilot study, tumor tissue, to comprehensively define individual inflammatory profiles. Additionally, while the long follow-up allows for accrual of many cases, misclassification of exposures that change over time due to temporal trends (e.g., medications) or increasing prevalence with age (e.g., chronic diseases) is a concern. To address this, we propose to collect updated exposure data from 15 studies with follow-up questionnaires. This collaborative study has substantial potential to further understanding of ovarian cancer, leading to improved prevention, via the following aims:

1. To assess the relationship of circulating levels of CRP, IL-6, and TNFα-R2 as well as the genetically-determined component of each marker (via Mendelian randomization analysis) with risk of ovarian cancer.
    a. We hypothesize that CRP, IL-6 and TNFα-R2, are positively associated with risk overall, that the association persists for at least 10 years after blood draw, and that the associations are stronger for aggressive tumor phenotypes and overweight/obese women.
    b. We hypothesize that considering patterns of CRP, IL-6 and TNFα-R2 levels (e.g., high levels of all three markers) will elucidate individuals at high risk of ovarian cancer.
    c. Based on biologic data, we hypothesize that genetically-determined levels of IL-6 and TNFα-R2, but not CRP, are associated with ovarian cancer risk, particularly for aggressive tumors.
2. To examine inflammation-related exposures with ovarian cancer risk overall and by subtype.
    a. We hypothesize that adiposity, inflammatory diet score, talc use, short or long sleep duration, IUD use, lifetime ovulatory cycles, allergies and asthma, autoimmune disease, cardiovascular disease, diabetes, and depression are associated with increased risk of ovarian cancer, while use of NSAIDS, antibiotics, statins, and bisphosphonates lower risk, with stronger associations for aggressive tumor phenotypes.

b. We hypothesize that grouping exposures based on associations with CRP, IL-6 and TNFα-R2 levels, and preliminarily by type of immune response elicited (Th1, Th2, Th17), will elucidate biologic mechanisms that are important in ovarian cancer pathogenesis.

c. Secondarily, we hypothesize that the inflammatory exposures in Aim 2a are more strongly related to high-grade serous tumors or tumors that have tumor-associated macrophages as assessed by IHC.

References

1. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer.* 2015;136(5):E359-386.

2. *Cancer Facts & Figures 2015.* Atlanta: American Cancer Society;2015.

3. Permuth-Wey J, Sellers TA. Epidemiology of ovarian cancer. *Methods Mol Biol.* 2009;472:413-437.

4. 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.

5. 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.

6. Poole EM, Merritt MA, Jordan SJ, et al. Hormonal and reproductive risk factors for epithelial ovarian cancer by tumor aggressiveness. *Cancer Epidemiol Biomarkers Prev.* 2013;22(3):429-437.

7. 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.

8. Fleming JS, Beaugie CR, Haviv I, Chenevix-Trench G, Tan OL. Incessant ovulation, inflammation and epithelial ovarian carcinogenesis: revisiting old hypotheses. *Mol Cell Endocrinol.* 2006;247(1-2):4-21.

9. Maccio A, Lai P, Santona MC, Pagliara L, Melis GB, Mantovani G. High serum levels of soluble IL-2 receptor, cytokines, and C reactive protein correlate with impairment of T cell response in patients with advanced epithelial ovarian cancer. *Gynecol Oncol.* 1998;69(3):248-252.

10. Moradi MM, Carson LF, Weinberg B, Haney AF, Twiggs LB, Ramakrishnan S. Serum and ascitic fluid levels of interleukin-1, interleukin-6, and tumor necrosis factor-alpha in patients with ovarian epithelial cancer. *Cancer.* 1993;72(8):2433-2440.

11. Maccio A, Madeddu C. Inflammation and ovarian cancer. *Cytokine.* 2012;58(2):133-147.

12. Rzymski P, Opala T, Wilczak M, Wozniak J, Sajdak S. Serum tumor necrosis factor alpha receptors p55/p75 ratio and ovarian cancer detection. *Int J Gynaecol Obstet.* 2005;88(3):292-298.

13. Lane D, Matte I, Rancourt C, Piche A. Prognostic significance of IL-6 and IL-8 ascites levels in ovarian cancer patients. *BMC Cancer.* 2011;11:210.

14. Lo CW, Chen MW, Hsiao M, et al. IL-6 trans-signaling in formation and progression of malignant ascites in ovarian cancer. *Cancer Res.* 2011;71(2):424-434.

15. Clendenen TV, Lundin E, Zeleniuch-Jacquotte A, et al. Circulating inflammation markers and risk of epithelial ovarian cancer. *Cancer Epidemiol Biomarkers Prev.* 2011;20(5):799-810.

16. 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.

17. Poole EM, Lee IM, Ridker PM, Buring JE, Hankinson SE, Tworoger SS. A prospective study of circulating C-reactive protein, interleukin-6, and tumor necrosis factor alpha receptor 2 levels and risk of ovarian cancer. *Am J Epidemiol.* 2013;178(8):1256-1264.

18. 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.

19. McSorley MA, Alberg AJ, Allen DS, et al. C-reactive protein concentrations and subsequent ovarian cancer risk. *Obstet Gynecol.* 2007;109(4):933-941.

20. Toriola AT, Grankvist K, Agborsangaya CB, Lukanova A, Lehtinen M, Surcel HM. Changes in pre-diagnostic serum C-reactive protein concentrations and ovarian cancer risk: a longitudinal study. *Ann Oncol.* 2011;22(8):1916-1921.

21. Lundin E, Dossus L, Clendenen T, et al. C-reactive protein and ovarian cancer: a prospective study nested in three cohorts (Sweden, USA, Italy). *Cancer Causes Control.* 2009;20(7):1151-1159.

22. Koenig W, Sund M, Frohlich M, et al. C-Reactive protein, a sensitive marker of inflammation, predicts future risk of coronary heart disease in initially healthy middle-aged men: results from the MONICA (Monitoring Trends and Determinants in Cardiovascular Disease) Augsburg Cohort Study, 1984 to 1992. *Circulation.* 1999;99(2):237-242.

23. Collaborative Group on Epidemiological Studies of Ovarian C, 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.

24. Autelitano DJ, Raineri L, Knight K, Bannister K, Rice GE. Performance of a multianalyte test as an aid for the diagnosis of ovarian cancer in symptomatic women. *J Transl Med.* 2012;10:45.

25. Sen S, Kuru O, Akbayir O, Oguz H, Yasasever V, Berkman S. Determination of serum CRP, VEGF, Leptin, CK-MB, CA-15-3 and IL-6 levels for malignancy prediction in adnexal masses. *J Turk Ger Gynecol Assoc.* 2011;12(4):214-219.