UC Irvine

Title

Risk Prediction for Epithelial Ovarian Cancer in 11 United States—Based Case-Control Studies: Incorporation of Epidemiologic Risk Factors and 17 Confirmed Genetic Loci

Permalink

https://escholarship.org/uc/item/4qd5v32h

Journal

American Journal of Epidemiology, 184(8)

ISSN

0002-9262

Authors

Clyde, Merlise A
Weber, Rachel Palmieri
Iversen, Edwin S
et al.

Publication Date

2016-10-15

DOI

10.1093/aje/kww091

Copyright Information

Peer reviewed

Original Contribution

Risk Prediction for Epithelial Ovarian Cancer in 11 United States–Based Case-Control Studies: Incorporation of Epidemiologic Risk Factors and 17 Confirmed Genetic Loci

Merlise A. Clyde, Rachel Palmieri Weber, Edwin S. Iversen, Elizabeth M. Poole, Jennifer A. Doherty, Marc T. Goodman, Roberta B. Ness, Harvey A. Risch, Mary Anne Rossing, Kathryn L. Terry, Nicolas Wentzensen, Alice S. Whittemore, Hoda Anton-Culver, Elisa V. Bandera, Andrew Berchuck, Michael E. Carney, Daniel W. Cramer, Julie M. Cunningham, Kara L. Cushing-Haugen, Robert P. Edwards, Brooke L. Fridley, Ellen L. Goode, Galina Lurie, Valerie McGuire, Francesmary Modugno, Kirsten B. Moysich, Sara H. Olson, Celeste Leigh Pearce, Malcolm C. Pike, Joseph H. Rothstein, Thomas A. Sellers, Weiva Sieh, Daniel Stram, Pamela J. Thompson, Robert A. Vierkant, Kristine G. Wicklund, Anna H. Wu, Argyrios Ziogas, Shelley S. Tworoger, and Joellen M. Schildkraut*, on behalf of the Ovarian Cancer Association Consortium

* Correspondence to Dr. Joellen M. Schildkraut, University of Virginia, Department of Public Health Sciences, PO Box 800765, 560 Ray C. Hunt Drive, Charlottesville, VA 22903 (e-mail: jms2yf@virginia.edu).

Previously developed models for predicting absolute risk of invasive epithelial ovarian cancer have included a limited number of risk factors and have had low discriminatory power (area under the receiver operating characteristic curve (AUC) < 0.60). Because of this, we developed and internally validated a relative risk prediction model that incorporates 17 established epidemiologic 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 (data accrued from 1992 to 2010). 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 that included age and study site only (AUC = 0.563). The best predictive power was obtained in the full model among women younger than 50 years of age (AUC = 0.714); however, the addition of SNPs increased the AUC the most for women older than 50 years of age (AUC = 0.638 vs. 0.616). Adapting this improved model to estimate absolute risk and evaluating it in prospective data sets is warranted.

genetic risk polymorphisms; model evaluation; ovarian cancer; risk model

Abbreviations: AUC, area under the curve; COGS, Collaborative Oncological Gene-Environment Study; EOC, epithelial ovarian cancer; GWAS, genome-wide association study; MCMC, Markov chain Monte Carlo; MHT, menopausal hormone therapy; OC, oral contraceptive; OCAC, Ovarian Cancer Association Consortium; ROC, receiver operating characteristics; SNP, single nucleotide polymorphism.

More than 21,000 cases of ovarian cancer and 14,180 deaths from ovarian cancer were expected in 2015, accounting for 5% of cancer deaths among women; most were expected to be cases of epithelial ovarian cancer (EOC) [1]. The 5-year survival rate for localized ovarian cancer is 92%, but most cases are diagnosed at a distant stage at which 5-year survival is only 27% [2]. 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 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). Risk prediction for EOC is challenging because of its rarity and the modest associations 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 EOC risk; family history of breast and/or ovarian cancers are positively associated with risk (8). Older age at menarche and use of menopausal hormone therapy (MHT) (particularly estrogen-only therapy) have been associated with a higher EOC risk, whereas breastfeeding and hysterectomy have been associated with a lower risk in some but not all studies (6, 9–16). Although results have been inconsistent, in a recent report of 12 population-based case-control studies, investigators 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 models generally have low discrimination (area under the receiver operator characteristic curve (AUC) < 0.60), which may be partly due to exclusion of women who report premenopausal hysterectomy (with or without unilateral oophorectomy), incomplete inclusion of risk factors (e.g., tubal ligation), or the specific subpopulations in which the model was evaluated (e.g., women having a hysterectomy or women with symptoms) (20–25). Although some existing risk models specifically address risk among carriers of the mutation in the breast cancer 1 and breast cancer 2 genes (BRCA1 and BRCA2) (26, 27), mutations are rare in the general population; prior models for women of average risk have not considered genetic susceptibility. Given the 17 confirmed genetic variants related to 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 factors and genetic data from 11 case-control studies in the United States 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 in which data were accrued from 1992–2010 (Table 1) (14, 35–45). All studies were population-based except for the Mayo Clinic Ovarian Cancer Case-Control Study (MAY), which was clinic-based; in that study, controls were women attending the Mayo Clinic's Departments of Family Medicine and General Internal Medicine for general medical examinations. 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 nonmelanoma skin cancer), and self-identified as white, non-Hispanic; most women were confirmed to be of European ancestry by genetic analysis. Control subjects had to have at least 1 intact ovary, and case patients had invasive EOC. Most case patients (81%) were recruited within 1 year of diagnosis. After exclusions, the analysis included data from 5,793 invasive EOC cases 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

Risk factors from each study, as well as demographic and clinical variables, were submitted to the OCAC data coordination center at Duke University, where common coding schemes were applied; data were originally collected via questionnaire. Data on the following risk factors were available in the majority of studies: age at menarche (continuous years); OC use (ever vs. 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 vs. never); duration of breastfeeding (continuous months); age at end of last pregnancy (continuous years); tubal ligation (yes vs. no); hysterectomy more than 1 year prior to diagnosis (cases) or interview/reference age (controls) (yes vs. no); endometriosis (yes vs. no); body mass index within 5 years of diagnosis/interview; menopause status at diagnosis (cases) or interview/reference age (controls) (premenopausal vs. postmenopausal); MHT use (ever vs. never); type of MHT (unopposed estrogen replacement therapy only vs. all other MHT use); history of breast cancer in a first-degree relative (yes vs. no); and history of ovarian cancer in a first-degree relative (yes vs. no). We considered additional potential risk factors (e.g., nonsteroidal antiinflammatory drug use, age at tubal ligation, age at menopause, and duration of MHT) that were ultimately not included because they were not significant predictors of EOC in preliminary models and were missing for a large percentage of participants. Because of frequency matching, age was included in all models to avoid bias (46), as were random effects for study sites.

Genetic susceptibility data

The OCAC evaluated 23,239 single nucleotide polymorphisms (SNPs) in 43 individual studies that were grouped into 34 case-control strata; 2 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 resulted in identification and confirmation of 17 susceptibility loci (Web Table 1, available at http://aje.oxfordjournals.org/) (28–34) that are

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, 1992–2010

First Author, Year (Reference No.)	Study Name	Study Acronym	Location	Period of Ascertainment	Median Age, years	Age Range, years	No. of Controls	No. of Cases	Response Rate, %[a]	
									Controls	Cases
Risch, 2006 (41)	Connecticut Ovarian Cancer Study	CON	CT	1998–2003	55	34–81	466	318	61	69
Rossing, 2007 (14)	Diseases of the Ovary and Their Evaluation	DOV	Western WA	2002–2009	57	35–74	1,527	894	62	74
Lurie, 2008 (38)	Hawaii Ovarian Cancer Case-Control Study	HAW	HI, southern CA	1993–2008	57	30–90	345	236	80	78
Lo-Ciganic, 2012 (37)	Novel Risk Factors and Potential Early Detection Markers for Ovarian Cancer	HOP	Western PA, northeast OH, western NY	2003–2009	57	30–94	1,561	570	68	71
Kelemen, 2008 (36)	Mayo Clinic Ovarian Cancer Case-Control Study	MAY	IA, IL, MN, ND, SD, WI	2000–2010	60	30–92	842	533	58	91
Schildkraut, 2010 (42)	North Carolina Ovarian Cancer Study	NCO	NC	1999–2008	57	30–75	751	651	60	67
Terry, 2005 (43)	New England Case-Control Study of Ovarian Cancer	NEC	NH, eastern MA	1992–2003	54	30–78	1,067	704	64	71
Bandera, 2011 (35)	New Jersey Ovarian Cancer Study	NJO	NJ	2002–2008	60	30–87	336	185	40	47
McGuire, 2004 (39)	Genetic Epidemiology of Ovarian Cancer Study	STA	San Francisco Bay Area, CA	1997–2001	50	30–65	330	276	75	75
Ziogas, 2000 (45)	University of California Irvine Ovarian Study	UCI	Southern CA	1993–2005	56	30–86	505	318	80	67
Pike, 2004 (40); Wu, 2009 (44)	Los Angeles County Case-Control Studies of Ovarian Cancer	USC	Los Angeles County, CA	1992–2002	57	30–85	1,782	1,108	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; NY, New York; OH, Ohio; PA, Pennsylvania; SD, South Dakota; WA, Washington; WI, Wisconsin.

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

included in our risk prediction model. Some, but not all, participants from the studies in our analysis contributed to the GWAS genotyping efforts (Mayo Clinic Ovarian Cancer Case-Control Study, North Carolina Ovarian Cancer Study (NCO), New England Case-Control Study of Ovarian Cancer (NEC)) and COGS (all studies except the Connecticut Ovarian Cancer Study (CON)), requiring imputation of missing SNPS for the remaining women.

Statistical analysis

We used generalized additive models (R package *mgcv;* R Foundation for Statistical Computing, Vienna, Austria) (47–49) with random effects for study site, fixed effects for categorical variables and SNPs, and smooth nonparametric functions for continuous variables for exploratory model fitting using subjects with complete data. Some evidence supports the idea that risk factor associations may vary by menopausal status (50). However, because age at menopause was missing for 59% of the postmenopausal women and is difficult to determine for some women because of premenopausal hysterectomy and hormone use, we fit separate models for women younger than 50 years of age and women 50 years or older. The generalized additive models suggested that nonlinear functions of the continuous variables could be approximated with linear functions of the variables ($P > 0.05$) except for duration of OC use. The square root of OC use duration did not produce a significant increase in the deviance compared with using the spline terms ($P = 0.2265$), and a linear term for OC use duration was rejected ($P = 0.0114$). We retained linear terms with the original continuous variables except for duration of OC use, for which we used the square root transformation. Nulliparity was included as a term for interaction with all variables that were not defined for nulliparous women (age at last pregnancy, breastfeeding, and breastfeeding duration).

Some data were missing for all risk factors except age; 80% of the participants were missing information on at least 1 risk factor (Table 2). Rather than limit our 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 were missing (in the case of data not missing at random) (52). Missing risk factors and indicators were modeled as functions of other risk covariates and of 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 (Vienna, Austria) (53). The increased sample size obtained by using participants with partial information can increase power, whereas 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) as

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = \alpha_{\text{site}_i}^g + \sum_{j=1}^{6} Z_{i,j}\beta_{j,c_i}^g + \sum_j X_{i,j}\beta_j^g \quad (1)$$

for the 2 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, as follows:

$$\alpha_{\text{site}_i}^g \overset{ind}{\sim} N(\mu_{\text{site}}, \sigma_{site}^2), \quad (2)$$

and random effects to account for birth cohort (c), as follows:

$$\beta_{j,c_i}^g \overset{ind}{\sim} N(\beta_j^g, \sigma_j^2) \quad (3)$$

for the 6 hormonally-related covariates Z (i.e., indicator of OC use, square root of OC use duration, indicator of MHT use, indicator of type of MHT use, 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 epidemiologic risk factors and the 17 SNPs). All of the group-specific means, β_j^g, for random-effects and fixed-effects coefficients for the other exposures were given independent normal prior distributions, with a mean β_j and a prior standard deviation of 1, which reflected the expectation that population log odds ratios 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 odds ratios from the GWAS and COGS samples independent from the 11 studies included in model development (Web 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 assuming Hardy-Weinberg equilibrium and a mass parameter in the Dirichlet equivalent to 1,000 observations; genotype probabilities were calculated using the minor allele frequencies estimated from GWAS and COGS samples from OCAC not used in this analysis (Web Table 2). Combined with genotype, 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 in 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 odds ratios were

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, From 11 Case-Control Studies, 1992–2010

Risk Factors Included in Model	Training Set						Evaluation Set					
	Controls (n = 7,586)			Cases (n = 4,662)			Controls (n = 1,926)			Cases (n = 1,131)		
	Mean (SD)	No.	%	Mean (SD)	No.	%	Mean (SD)	No.	%	Mean (SD)	No.	%
Age at diagnosis/interview, years	56.2 (11.6)			57.58 (10.9)			56.69 (11.7)			57.51 (10.9)		
Age at menarche, years	12.7 (1.6)			12.6 (1.5)			12.7 (1.5)			12.6 (1.5)		
Missing age		63	1		95	2		19	1		28	2
OC use												
Ever used		5,341	70		2,750	59		1,350	70		682	60
Missing OC use		69	1		58	1		12	1		16	1
Months of OC use	74.7 (69.4)	57[b]		58.3 (61.3)	36[b]		76.3 (70.9)	58[b]		59.1 (55.0)	48[b]	
Missing months of OC use		89	1		79	2		19	1		21	2
Pregnancy history												
No. of full-term pregnancies	2.2 (1.5)			1.9 (1.6)			2.2 (1.6)			1.9 (1.5)		
Missing no. of full-term pregnancies		44	1		31	1		8	<1		10	1
No. of pregnancies	3.2 (1.7)			3.0 (1.7)			3.2 (1.7)			2.9 (1.6)		
Missing no. of pregnancies		45	1		31	1		8	<1		10	1
Non–full-term pregnancies	0.65 (1.1)			0.52 (1.0)			0.60 (1.0)			0.53 (1.0)		
Missing no. of non–full-term pregnancies		45	1		31	1		8	<1		10	1
Age at end of last pregnancy, years	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		3,250	43		1,507	32		799	41		393	35
Missing breastfeeding status		1,201	16		621	13		306	16		128	11
Months of breastfeeding	14.2 (16.3)			11.6 (15.8)			14.7 (15.8)			10.8 (12.7)		
Missing breastfeeding duration		1,203	16		623	13		306	16		128	11
Tubal ligation												
Had tubal ligation		1,585	21		709	15		380	20		185	16
Missing information		892	12		329	7		232	12		70	6
Endometriosis												
Had endometriosis		585	8		475	10		137	7		124	11
Missing information		354	5		367	8		78	4		93	8
Family history (first-degree relative)												
Breast cancer		1,073	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
BMI[c]												
BMI	26.44 (6.11)			26.82 (6.42)			26.50 (6.09)			26.47 (6.12)		
Missing BMI		342	5		275	6		74			67	6
Aspirin use												
Irregular or no use		3,786	50		2,349	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		3,367	44		2,146	46		856	44		502	44

Table continues

Table 2. Continued

Risk Factors Included in Model	Training Set						Evaluation Set					
	Controls (n = 7,586)			Cases (n = 4,662)			Controls (n = 1,926)			Cases (n = 1,131)		
	Mean (SD)	No.	%	Mean (SD)	No.	%	Mean (SD)	No.	%	Mean (SD)	No.	%
Menopausal status												
Postmenopausal		4,818	64		3,215	69		1,247	65		774	68
Missing menopausal status		174	2		72	2		46	2		20	2
Hysterectomy												
Had hysterectomy[d]		1,015	13		738	16		248	13		167	15
Missing information		147	2		595	13		36	2		151	13
MHT												
Ever used MHT		2,938	39		1,907	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	11		152	13
Missing type of MHT		477	6		443	10		110	6		114	10
rs1243180[e]												
No minor alleles		2,770	37		1,505	32		628	33		396	35
1 minor allele		2,313	30		1,512	32		631	33		342	30
2 minor alleles		523	7		368	8		140	7		86	8
rs2072590[e]												
No minor alleles		2,652	35		1,451	31		620	32		364	32
1 minor allele		2,414	32		1,533	33		649	34		355	31
2 minor alleles		546	7		404	9		132	7		106	9
rs11782652[e]												
No minor alleles		4,839	64		2,890	62		1,229	64		693	61
1 minor allele		734	10		476	10		163	8		125	11
2 minor alleles		25	<1		19	<1		6	<1		5	<1
rs10088218[e]												
No minor alleles		4,198	55		2,656	57		1,032	54		630	56
1 minor allele		1,306	17		689	15		348	18		185	16
2 minor alleles		105	1		43	1		21	1		9	1
rs757210[e]												
No minor alleles		2,230	29		1,292	28		555	29		321	28
1 minor allele		2,599	34		1,567	34		662	34		379	34
2 minor alleles		762	10		525	11		180	9		123	11
rs9303542[e]												
No minor alleles		2,982	39		1,628	35		691	36		423	37
1 minor allele		2,219	29		1,456	31		598	31		337	30
2 minor alleles		407	5		301	6		110	6		65	6
rs7651446[e]												
No minor alleles		5,070	67		2,952	63		1,273	66		699	62
1 minor allele		527	7		423	9		121	6		117	10
2 minor alleles		15	<1		13	<1		7	<1		9	1
rs3814113[e]												
No minor alleles		2,597	34		1,721	37		643	33		437	39
1 minor allele		2,421	32		1,377	30		623	32		318	28
2 minor alleles		594	8		290	6		135	7		70	6
rs8170[e]												
No minor alleles		3,703	49		2,192	47		949	49		510	45
1 minor allele		1,735	23		1,077	23		414	21		284	25
2 minor alleles		174	2		119	3		38	2		31	3

Table continues

Table 2. Continued

Risk Factors Included in Model	Training Set						Evaluation Set					
	Controls (n = 7,586)			Cases (n = 4,662)			Controls (n = 1,926)			Cases (n = 1,131)		
	Mean (SD)	No.	%	Mean (SD)	No.	%	Mean (SD)	No.	%	Mean (SD)	No.	%
rs10069690[e]												
No minor alleles		3,061	40		1,757	38		765	40		441	39
1 minor allele		2,147	28		1,350	29		523	27		322	28
2 minor alleles		351	5		234	5		101	5		58	5
rs56318008[e]												
No minor alleles		4,152	55		2,385	51		1,028	53		584	52
1 minor allele		1,353	18		915	20		348	18		221	20
2 minor alleles		106	1		87	2		25	1		20	2
rs58722170[e]												
No minor alleles		3,403	45		2,029	44		832	43		462	41
1 minor allele		1,941	26		1,201	26		503	26		318	28
2 minor alleles		267	4		157	4		66	3		45	4
rs17329882[e]												
No minor alleles		3,317	44		1,874	40		836	43		477	42
1 minor allele		1,989	26		1,302	28		491	25		292	26
2 minor alleles		305	4		211	5		74	4		56	5
rs116133110[e]												
No minor alleles		2,702	36		1,678	36		626	33		411	36
1 minor allele		2,337	31		1,419	30		634	33		346	31
2 minor alleles		572	8		290	6		141	7		68	6
rs635634[e]												
No minor alleles		3,597	47		2,074	44		895	46		497	44
1 minor allele		1,803	24		1,176	25		448	23		291	26
2 minor alleles		211	3		137	3		58	3		37	3
chr17_29181220[e]												
No minor alleles		2,916	38		1,845	40		716	37		461	41
1 minor allele		2,241	30		1,338	29		562	29		308	27
2 minor alleles		454	6		204	4		123	6		56	5
rs183211[e]												
No minor alleles		3,241	43		1,859	40		824	43		447	40
1 minor allele		2,051	27		1,290	28		488	25		332	29
2 minor alleles		319	4		238	5		89	5		46	4

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

[a] Frequency distributions are based on nonmissing data. Percent missing is based on the variable of interest and any upper level variable related to it. For example, women who are missing information on 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] Median months of OC use.

[c] Weight (kg)/height (m)^2.

[d] Women who reported hysterectomies more than 1 year prior to diagnosis (cases) or interview/reference date (controls) were considered to have had a hysterectomy.

[e] Missing genotype data were approximately the same across the 17 single nucleotide polymorphisms. The percentages of participants missing genotype data were as follows: 26% for training set controls, 27%–28% for training set cases and evaluation set controls, and 27% for evaluation set cases.

estimated by the median of the samples from the posterior distribution of each of the parameters; Bayesian 95% confidence intervals 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 1. For comparison, we also fit a model that was adjusted for study site and age only (baseline model) and one that was adjusted for study site, age, and SNPs, omitting epidemiologic risk factors.

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, From 11 Case-Control Studies, 1992–2010[a]

Risk Factor	Covariates Included in Imputation Model for Risk Factor	Distribution
SNP genotypes	Site	Multinomial-Dirichlet
Family history of ovarian cancer	Site	Bernoulli
Family history of breast cancer	Family history of ovarian cancer and site	Bernoulli
Endometriosis	Cohort, age, and site	Bernoulli
Menopausal status	Alcohol, smoking status, age, and site	Bernoulli
Tubal ligation	Endometriosis, educational level, age, cohort, and site	Bernoulli
Hysterectomy	Endometriosis, tubal ligation, family history of breast cancer, family history of ovarian cancer, age, cohort, and site	Bernoulli
Height (BMI)	Site and cohort	Gaussian
Weight (BMI)	Site, cohort, height, age, smoking status, and educational level	Gaussian
Aspirin use	Site, cohort, age, smoking status, and BMI	Bernoulli
Ever used MHT	Menopausal status, hysterectomy, educational level, age, cohort, and site	Bernoulli
Type of MHT	Ever used MHT, menopausal status, hysterectomy, educational level, age, cohort, and site	Bernoulli
Age at menarche	Age, cohort, and site	Truncated Student t
Ever used OCs	Cohort and site	Bernoulli
Duration of OC use	Ever used OCs, age, cohort, and site	Truncated Gaussian
No. of pregnancies	Hysterectomy, tubal ligation, ever used OCs, endometriosis, educational level, smoking, alcohol use, age, cohort, and site	Poisson
No. of full-term births	No. of pregnancies, hysterectomy, tubal ligation, ever used OCs, endometriosis, educational level, smoking, alcohol use, age, cohort, and site	Binomial
Age at end of last pregnancy	No. of pregnancies, age at menarche, smoking status, educational level, age, cohort, and site	Truncated Gaussian
Ever breastfed	No. of pregnancies, smoking status, educational level, cohort, and site	Bernoulli
Duration of breastfeeding	No. of pregnancies, smoking status, educational level, age, cohort, and 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 Covariates column.

Model validation

We compared the models with and without SNPs and with and without the epidemiologic variables (all models included reference age and study) based on 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). The average age was 57 years. In both the training and evaluation sets, case patients were less likely to use OCs, have been pregnant, and have had a tubal ligation than were controls and were more likely to have a family history of breast or ovarian cancer and to use MHT. The distribution of SNPs was similar to those observed in the GWAS and COGs data sets.

Table 4 provides estimates of the log odds ratios (medians) and 95% Bayesian confidence intervals 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 (Web Table 3). Most risk factors included in the model were statistically significant predictors among women younger than 50 years of age; however, in general, the directions of associations were comparable across groups. Notably, some associations were weaker among older women than among younger women, including duration of OC use, number of pregnancies, breastfeeding, family history of breast or ovarian cancers, endometriosis, tubal ligation, MHT use and type, and hysterectomy, whereas low-dose aspirin use showed a significant inverse association in women 50 years of age or older. Furthermore, more SNPs were significant for women 50 years or older, who comprised the majority of women in this study. Endometriosis, duration of OC use, tubal ligation, family history of breast or ovarian cancer, number of non–full-term pregnancies, and SNPS rs2072590, rs10088218 in 8q24, rs9303542, rs7651446 in 5p15,

Table 4. Estimates of Log Odds Ratios and 95% Bayesian Confidence Intervals for Risk Factors Included in the Invasive Epithelial Ovarian Cancer Relative Risk Prediction Model Containing 17 Confirmed Single Nucleotide Polymorphisms, Stratified by Age at Diagnosis (Cases) or Interview/Reference Age (Controls), From 11 Case-Control Studies, 1992–2010[a]

Risk Factor	Age at Diagnosis/Interview, years			
	<50 (n = 1,286 Cases and 2,473 Controls)		≥50 (n = 3,376 Cases and 5,113 Controls)	
	Median	95% CI	Median	95% CI
Age	0.0308	0.0117, 0.0438	−0.0067	−0.0205, 0.0014
High-dose aspirin use	0.05	−0.4624, 0.6254	−0.1223	−0.3517, 0.062
Low-dose aspirin use	−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
Ever breastfed	−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 of breast cancer	0.317	0.0885, 0.5534	0.1663	0.0537, 0.2902
Family history of ovarian cancer	1.3687	0.9383, 1.7791	0.4949	0.2625, 0.7273
Hysterectomy and no MHT use	−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 with estrogen and no hysterectomy	1.5661	0.992, 1.8842	−0.1107	−0.3277, 0.1101
MHT with estrogen and hysterectomy	−2.1774	−2.7231, −1.5081	0.2408	−0.027, 0.4781
Other MHT without hysterectomy	0.1682	−0.2312, 0.482	−0.182	−0.3235, −0.0267
Other MHT 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 of 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; OC, oral contraceptive.

[a] Estimates and intervals are based on the training set only.

rs3814113, rs56318008, and rs183211 contributed significantly to all of the group-specific models.

The AUCs for models for all women, women younger than 50 years, and women 50 years and older both without and with SNPs included are shown in Figure 1A and 1B, respectively. The inclusion of the SNPs provided a small improvement (change in the AUC = 0.015) in predictions for the validation data in terms of AUC for all women, with the biggest improvement for women 50 years of age and older (0.026 increase). Among all women, the AUC was 0.664 in the model with SNPs and 0.649 in the model without SNPs (but including epidemiologic factors), which is a marked improvement over the AUC for the models with age and study site alone (AUC = 0.563) and those with 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 epidemiologic factors is better than the AUC for the model with age, study site, and SNPs alone was 99.8%, whereas there was a 70% chance that the addition of SNPs improved AUC over the model with age, study site, and epidemiologic factors. The best predictive power was obtained for women younger than 50 years: The AUCs were 0.714 and 0.713 in the models with and without the SNPs, respectively. Lower AUCs were observed in women 50 years of age or older (with SNPs, AUC = 0.638; without SNPs, AUC = 0.612). Finally, 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 validated SNPS for EOC) until the

AUC exceeded 0.75. Under this setting, on average 58 additional SNPS would be needed (95% confidence interval: 39, 79) to increase the AUC from 0.66 to 0.75. Figure 2 and Web 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 did 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 larger sample size provided ample power for stratification by age and permitted us to include a much larger number of accepted epidemiologic risk factors, as well as 17 genetic loci. This and 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, ours was a joint model for disease status, risk factors, and missingness. A strength
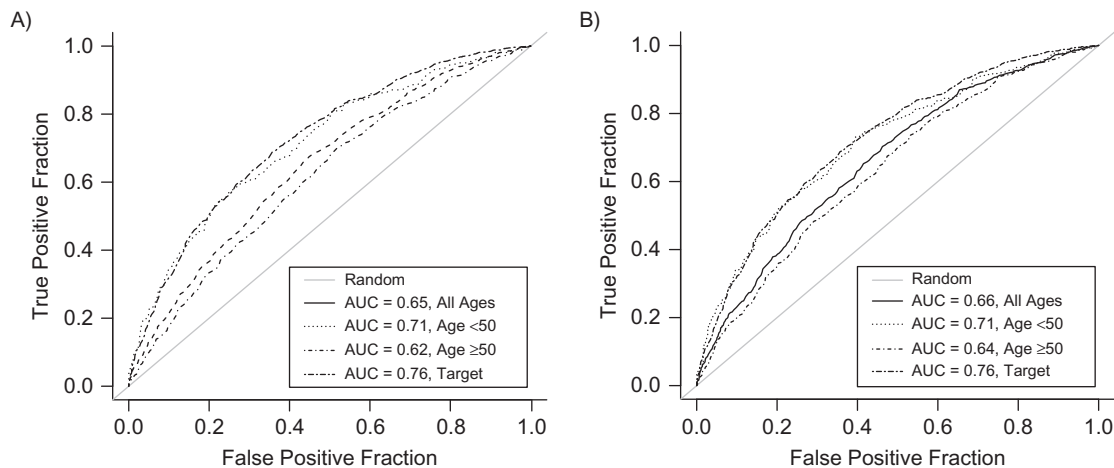
Figure 1. Invasive epithelial ovarian cancer relative risk prediction model from the Ovarian Cancer Association Consortium, 1992–2010. Receiver operating characteristic curve for models A) without and B) with single nucleotide polymorphisms (SNPs). The receiver operating characteristic (ROC) curve plots 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 age: all ages, women younger than 50 years of age, and women 50 years of age or older. The area under the curve, 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 is depicted based on adding additional hypothetical SNPs with a minor allele frequency of 0.20 and log odds ratio of 0.15 (similar to the current data) until the area under the curve is 0.75 or more; on average, 58 additional SNPs would be needed (95% confidence interval: 39, 79).

Table 5. Predictive Power for Relative Risk Prediction Models for Invasive Epithelial Ovarian Cancer That Include Age, Study Site, 17 Epidemiologic Risk Factors, or 17 Confirmed Genetic Susceptibility Variants, From 11 Case-Control Studies, 1992–2010

Age	Study Site	Epidemiologic Risk Factors	SNPs	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: AUC, area under the receiver operating characteristic curve; SNPs, single nucleotide polymorphisms.

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 lead to biased inferences, 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, for which 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-

Figure 2. Calibration plots for risk scores from the invasive epithelial ovarian cancer relative risk prediction model from the Ovarian Cancer Association Consortium, 1992–2010. 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 10 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.

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 older than 50 years of age, 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 1 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 women (AUC = 0.714) than for 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 younger than 50 years of age. One reason for the improved prediction in younger women is that many of the risk factors occur during premenopause 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 those from studies of individual risk factors that suggested, for example, that the inverse association with 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 that incorporated epidemiologic risk factors by a small amount (with SNPs, AUC = 0.664; without SNPs, AUC = 0.649). Our model pertains to women of average baseline risk, and mutation status of highly penetrant susceptibility genes such as *BRCA1* and *BRCA2* was not included because 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 may only modestly improve upon a risk model's discriminatory accuracy (59), and a very large number of susceptibility SNPs are required to make a substantial impact because of their small relative risks (60). Our simulation results suggest that an additional 39–79 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, for which a 3–4 unit increase can be achieved with addition of 60–70 SNPs (56, 58, 61–64).

The model may be improved by extension to predict histologic subtypes of EOC, because 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 ([65], [66]). Finally, these results may not be generalizable to other racial/ethnic groups or to women in 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.

ACKNOWLEDGEMENTS

REFERENCES

1. American Cancer Society. *Cancer Facts & Figures 2015*. Atlanta, GA: American Cancer Society; 2015. http://www.cancer.org/acs/groups/content/@editorial/documents/document/acspc-044552.pdf. Accessed September 14, 2016.
2. Siegel R, Ma J, Zou Z, et al. Cancer statistics, 2014. *CA Cancer J Clin*. 2014;64(1):9–29.
3. Freedman AN, Seminara D, Gail MH, et al. Cancer risk prediction models: a workshop on development, evaluation, and application. *J Natl Cancer Inst*. 2005;97(10):715–723.
4. Collaborative Group on Epidemiological Studies of Ovarian Cancer, Beral V, Doll R, et al. 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): 303–314.
5. Adami HO, Hsieh CC, Lambe M, et al. Parity, age at first childbirth, and risk of ovarian cancer. *Lancet*. 1994; 344(8932):1250–1254.
6. Rice MS, Murphy MA, Tworoger SS. Tubal ligation, hysterectomy and ovarian cancer: a meta-analysis. *J Ovarian Res*. 2012;5(1):13.
7. Sieh W, Salvador S, McGuire V, et al. Tubal ligation and risk of ovarian cancer subtypes: a pooled analysis of case-control studies. *Int J Epidemiol*. 2013;42(2):579–589.
8. Schildkraut JM, Risch N, Thompson WD. Evaluating genetic association among ovarian, breast, and endometrial cancer: evidence for a breast/ovarian cancer relationship. *Am J Hum Genet*. 1989;45(4):521–529.
9. Beral V, Million Women Study Collaborators, Bull D, et al. Ovarian cancer and hormone replacement therapy in the Million Women Study. *Lancet*. 2007;369(9574):1703–1710.
10. Coughlin SS, Giustozzi A, Smith SJ, et al. A meta-analysis of estrogen replacement therapy and risk of epithelial ovarian cancer. *J Clin Epidemiol*. 2000;53(4):367–375.
11. Gong TT, Wu QJ, Vogtmann E, et al. Age at menarche and risk of ovarian cancer: a meta-analysis of epidemiological studies. *Int J Cancer*. 2013;132(12):2894–2900.
12. Luan NN, Wu QJ, Gong TT, et al. Breastfeeding and ovarian cancer risk: a meta-analysis of epidemiologic studies. *Am J Clin Nutr*. 2013;98(4):1020–1031.
13. Pearce CL, Chung K, Pike MC, et al. Increased ovarian cancer risk associated with menopausal estrogen therapy is reduced by adding a progestin. *Cancer*. 2009;115(3): 531–539.
14. Rossing MA, Cushing-Haugen KL, Wicklund KG, et al. Menopausal hormone therapy and risk of epithelial ovarian cancer. *Cancer Epidemiol Biomarkers Prev*. 2007;16(12): 2548–2556.
15. Trabert B, Wentzensen N, Yang HP, et al. Ovarian cancer and menopausal hormone therapy in the NIH-AARP Diet and Health Study. *Br J Cancer*. 2012;107(7):1181–1187.
16. Zhou B, Sun Q, Cong R, et al. Hormone replacement therapy and ovarian cancer risk: a meta-analysis. *Gynecol Oncol*. 2008;108(3):641–651.
17. Trabert B, Ness RB, Lo-Ciganic WH, 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;106(2):djt431.
18. Heidemann LN, Hartwell D, Heidemann CH, et al. The relation between endometriosis and ovarian cancer – a review. *Acta Obstet Gynecol Scand*. 2014;93(1):20–31.
19. Pearce CL, Templeman C, Rossing MA, et al. Association between endometriosis and risk of histological subtypes of ovarian cancer: a pooled analysis of case-control studies. *Lancet Oncol*. 2012;13(4):385–394.
20. Andersen MR, Goff BA, Lowe KA, et al. Use of a Symptom Index, CA125, and HE4 to predict ovarian cancer. *Gynecol Oncol*. 2010;116(3):378–383.
21. Collins GS, Altman DG. Identifying women with undetected ovarian cancer: independent and external validation of QCancer(®) (Ovarian) prediction model. *Eur J Cancer Care (Engl)*. 2013;22(4):423–429.
22. Hartge P, Whittemore AS, Itnyre J, et al. Rates and risks of ovarian cancer in subgroups of white women in the United States. The Collaborative Ovarian Cancer Group. *Obstet Gynecol*. 1994;84(5):760–764.
23. Pfeiffer RM, Park Y, Kreimer AR, 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.
24. Rosner BA, Colditz GA, Webb PM, et al. Mathematical models of ovarian cancer incidence. *Epidemiology*. 2005; 16(4):508–515.

25. Vitonis AF, Titus-Ernstoff L, Cramer DW. Assessing ovarian cancer risk when considering elective oophorectomy at the time of hysterectomy. *Obstet Gynecol.* 2011;117(5):1042–1050.

26. Chen S, Iversen ES, Friebel T, et al. Characterization of BRCA1 and BRCA2 mutations in a large United States sample. *J Clin Oncol.* 2006;24(6):863–871.

27. Lee AJ, Cunningham AP, Kuchenbaecker KB, et al. BOADICEA breast cancer risk prediction model: updates to cancer incidences, tumour pathology and web interface. *Br J Cancer.* 2014;110(2):535–545.

28. Bojesen SE, Pooley KA, Johnatty SE, et al. Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. *Nat Genet.* 2013;45(4):371–384, 384e1–384e2.

29. Bolton KL, Tyrer J, Song H, et al. Common variants at 19p13 are associated with susceptibility to ovarian cancer. *Nat Genet.* 2010;42(10):880–884.

30. Goode EL, Chenevix-Trench G, Song H, et al. A genome-wide association study identifies susceptibility loci for ovarian cancer at 2q31 and 8q24. *Nat Genet.* 2010;42(10):874–879.

31. Permuth-Wey J, Lawrenson K, Shen HC, et al. Identification and molecular characterization of a new ovarian cancer susceptibility locus at 17q21.31. *Nat Commun.* 2013;4:1627.

32. Pharoah PD, Tsai YY, Ramus SJ, et al. GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. *Nat Genet.* 2013;45(4):362–370.

33. Song H, Ramus SJ, Tyrer J, et al. A genome-wide association study identifies a new ovarian cancer susceptibility locus on 9p22.2. *Nat Genet.* 2009;41(9):996–1000.

34. Kuchenbaecker KB, Ramus SJ, Tyrer J, et al. Identification of six new susceptibility loci for invasive epithelial ovarian cancer. *Nat Genet.* 2015;47(2):164–171.

35. Bandera EV, King M, Chandran U, et al. Phytoestrogen consumption from foods and supplements and epithelial ovarian cancer risk: a population-based case control study. *BMC Womens Health.* 2011;11:40.

36. Kelemen LE, Sellers TA, Schildkraut JM, et al. Genetic variation in the one-carbon transfer pathway and ovarian cancer risk. *Cancer Res.* 2008;68(7):2498–2506.

37. Lo-Ciganic WH, Zgibor JC, Bunker CH, et al. Aspirin, nonaspirin nonsteroidal anti-inflammatory drugs, or acetaminophen and risk of ovarian cancer. *Epidemiology.* 2012;23(2):311–319.

38. Lurie G, Wilkens LR, Thompson PJ, et al. Combined oral contraceptive use and epithelial ovarian cancer risk: time-related effects. *Epidemiology.* 2008;19(2):237–243.

39. McGuire V, Felberg A, Mills M, et al. Relation of contraceptive and reproductive history to ovarian cancer risk in carriers and noncarriers of BRCA1 gene mutations. *Am J Epidemiol.* 2004;160(7):613–618.

40. Pike MC, Pearce CL, Peters R, et al. Hormonal factors and the risk of invasive ovarian cancer: a population-based case-control study. *Fertil Steril.* 2004;82(1):186–195.

41. Risch HA, Bale AE, Beck PA, et al. PGR +331 A/G and increased risk of epithelial ovarian cancer. *Cancer Epidemiol Biomarkers Prev.* 2006;15(9):1738–1741.

42. Schildkraut JM, Iversen ES, Wilson MA, 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, et al. Androgen receptor cytosine, adenine, guanine repeats, and haplotypes in relation to ovarian cancer risk. *Cancer Res.* 2005;65(13):5974–5981.

44. Wu AH, Pearce CL, Tseng CC, et al. Markers of inflammation and risk of ovarian cancer in Los Angeles County. *Int J Cancer.* 2009;124(6):1409–1415.

45. Ziogas A, Gildea M, Cohen P, et al. Cancer risk estimates for family members of a population-based family registry for breast and ovarian cancer. *Cancer Epidemiol Biomarkers Prev.* 2000;9(1):103–111.

46. Janes H, Pepe MS. Adjusting for covariates in studies of diagnostic, screening, or prognostic markers: an old concept in a new setting. *Am J Epidemiol.* 2008;168(1):89–97.

47. R Core Team. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing; 2014. http://www.R-project.org/.

48. Wood SN. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J R Stat Soc Series B Stat Methodol.* 2011;73(1):3–36.

49. Wood SN. *Generalized Additive Models: An Introduction with R.* Boca Raton, FL: Chapman and Hall/CRC; 2006.

50. Moorman PG, Calingaert B, Palmieri RT, et al. Hormonal risk factors for ovarian cancer in premenopausal and postmenopausal women. *Am J Epidemiol.* 2008;167(9):1059–1069.

51. Gelman A, Hill J. *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge, New York: Cambridge University Press; 2007.

52. Little RJA, Rubin DB. *Statistical Analysis With Missing Data.* Hoboken, NJ: 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). Vienna, Austria, March 20–22, 2003.

54. Greenland S, Finkle WD. A critical look at methods for handling missing covariates in epidemiologic regression analyses. *Am J Epidemiol.* 1995;142(12):1255–1264.

55. Hoff PD. *A First Course in Bayesian Statistical Methods.* New York, NY: Springer; 2009.

56. Hüsing A, Canzian F, Beckmann L, et al. Prediction of breast cancer risk by genetic risk factors, overall and by hormone receptor status. *J Med Genet.* 2012;49(9):601–608.

57. Park JH, Gail MH, Greene MH, et al. Potential usefulness of single nucleotide polymorphisms to identify persons at high cancer risk: an evaluation of seven common cancers. *J Clin Oncol.* 2012;30(17):2157–2162.

58. Wacholder S, Hartge P, Prentice R, et al. Performance of common genetic variants in breast-cancer risk models. *N Engl J Med.* 2010;362(11):986–993.

59. Pepe MS, Janes H, Longton G, et al. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *Am J Epidemiol.* 2004;159(9):882–890.

60. Gail MH. Discriminatory accuracy from single-nucleotide polymorphisms in models to predict breast cancer risk. *J Natl Cancer Inst.* 2008;100(14):1037–1041.

61. Darabi H, Czene K, Zhao W, et al. Breast cancer risk prediction and individualised screening based on common genetic variation and breast density measurement. *Breast Cancer Res.* 2012;14(1):R25.

62. Mealiffe ME, Stokowski RP, Rhees BK, et al. Assessment of clinical validity of a breast cancer risk model combining

genetic and clinical information. *J Natl Cancer Inst*. 2010; 102(21):1618–1627.

63. Pharoah PD, Antoniou AC, Easton DF, et al. Polygenes, risk prediction, and targeted prevention of breast cancer. *N Engl J Med*. 2008;358(26):2796–2803.

64. Gail MH. Value of adding single-nucleotide polymorphism genotypes to a breast cancer risk model. *J Natl Cancer Inst*. 2009;101(13):959–963.

65. Tice JA, Cummings SR, Smith-Bindman R, et al. Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model. *Ann Intern Med*. 2008;148(5):337–347.

66. Tworoger SS, Zhang X, Eliassen AH, et al. Inclusion of endogenous hormone levels in risk prediction models of postmenopausal breast cancer. *J Clin Oncol*. 2014;56: 1068.