

Keywords: ovarian cancer; risk model; reproductive risk factors

An epidemiologic risk prediction model for ovarian cancer in Europe: the EPIC study

K Li^{1,32}, A Hüsing^{1,32}, R T Fortner¹, A Tjønneland², L Hansen², L Dossus^{3,4}, J Chang-Claude¹, M Bergmann⁵, A Steffen⁵, C Bamia⁶, D Trichopoulos^{7,8,9}, A Trichopoulou^{8,9}, D Palli¹⁰, A Mattiello¹¹, C Agnoli¹², R Tumino¹³, N C Onland-Moret¹⁴, P H Peeters^{14,15}, H B(as) Bueno-de-Mesquita¹⁶, I T Gram¹⁷, E Weiderpass^{17,18,19}, E Sánchez-Cantalejo^{20,21}, M-D Chirlaque^{21,22}, E J Duell²³, E Ardanaz^{21,24}, A Idahl²⁵, E Lundin²⁶, K-T Khaw²⁷, R C Travis²⁸, M A Merritt²⁹, M J Gunter²⁹, E Riboli²⁹, P Ferrari³⁰, K Terry^{7,31}, D Cramer^{7,31} and R Kaaks¹

¹Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany; ²The Danish Cancer Society Research Center, Copenhagen, Denmark; ³Inserm, Centre for research in Epidemiology and Population Health (CESP), U1018, Nutrition, Hormones and Women's Health team, F-94805 Villejuif, France; ⁴University Paris Sud, UMRS 1018, F-94805 Villejuif, France; ⁵German Institute of Human Nutrition in Potsdam-Rehbruecke, Potsdam, Germany; ⁶Department of Hygiene, Epidemiology and Medical Statistics, University of Athens Medical School, Athens, Greece; ⁷Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA; ⁸Bureau of Epidemiologic Research, Academy of Athens, Athens, Greece; ⁹Hellenic Health Foundation, Athens, Greece; ¹⁰Molecular and Nutritional Epidemiology Unit, Cancer Research and Prevention Institute—ISPO, Florence, Italy; ¹¹Dipartimento di Medicina Clinica e Chirurgia, University of Naples Federico II, Naples, Italy; ¹²Epidemiology and Prevention Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy; ¹³Cancer Registry and Histopathology Unit, 'Civic—M.P. Arezzo' Hospital, Ragusa, Italy; ¹⁴Department of Epidemiology, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands; ¹⁵MRC-PHE Centre for Environment and Health, Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK; ¹⁶Department for Determinants of Chronic Diseases (DCD), National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands; ¹⁷Department of Community Medicine, Faculty of Health Sciences, The Arctic University of Norway, Tromsø, Norway; ¹⁸Department of Research, Cancer Registry of Norway, Oslo, Norway; ¹⁹Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; ²⁰Escuela Andaluza de Salud Pública, Instituto de Investigación Biosanitaria de Granada (Granada.ibs), Granada, Spain; ²¹CIBER de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain; ²²Department of Epidemiology, Murcia Regional Health Authority, Murcia, Spain; ²³Unit of Nutrition, Environment and Cancer Epidemiology Research Program, Bellvitge Biomedical Research Institute (IDIBELL), Catalan Institute of Oncology (ICO), Barcelona, Spain; ²⁴Navarre Public Health Institute, Pamplona, Spain; ²⁵Department of Clinical Sciences, Obstetrics and Gynecology and Department of Public Health and Clinical Medicine, Nutritional Research Umeå University, Umeå, Sweden; ²⁶Department of Medical Biosciences, Pathology Umeå University, Umeå, Sweden; ²⁷University of Cambridge, School of Clinical Medicine, Cambridge, UK; ²⁸Cancer Epidemiology Unit, Nuffield Department of Population Health University of Oxford, Oxford, UK; ²⁹Department of Epidemiology and Biostatistics, School of Public Health, Imperial College, London, UK; ³⁰International Agency for Research on Cancer, Lyon, France and ³¹Obstetrics and Gynecology Epidemiology Center, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

Background: Ovarian cancer has a high case-fatality ratio, largely due to late diagnosis. Epidemiologic risk prediction models could help identify women at increased risk who may benefit from targeted prevention measures, such as screening or chemopreventive agents.

Methods: We built an ovarian cancer risk prediction model with epidemiologic risk factors from 202 206 women in the European Prospective Investigation into Cancer and Nutrition study.

Results: Older age at menopause, longer duration of hormone replacement therapy, and higher body mass index were included as increasing ovarian cancer risk, whereas unilateral ovariectomy, longer duration of oral contraceptive use, and higher number of full-term pregnancies were decreasing risk. The discriminatory power (overall concordance index) of this model, as examined with five-fold cross-validation, was 0.64 (95% confidence interval (CI): 0.57, 0.70). The ratio of the expected to observed number of ovarian cancer cases occurring in the first 5 years of follow-up was 0.90 (293 out of 324, 95% CI: 0.81–1.01), in general there was no evidence for miscalibration.

Conclusion: Our ovarian cancer risk model containing only epidemiological data showed modest discriminatory power for a Western European population. Future studies should consider adding informative biomarkers to possibly improve the predictive ability of the model.

*Correspondence: Dr K Li; E-mail: k.li@dkfz.de

³²These authors contributed equally to this work

Received 29 September 2014; revised 22 December 2014; accepted 29 December 2014; published online 5 March 2015

© 2015 Cancer Research UK. All rights reserved 0007–0920/15

The age-standardised incidence rate of ovarian cancer in Europe is ~13 per 100 000, but its high fatality makes it the fifth leading cause of cancer deaths among women in this region (Ferlay *et al*, 2013). Among European women aged 55–74 years, the 5-year relative survival rate for ovarian cancer overall is 37.1% (Oberaigner *et al*, 2012), which, however, would increase up to 90% if the tumours were detected at the localised stages (American Cancer Society, 2014), implying the importance of early detection in reducing the ovarian cancer mortality rate.

In epidemiological literature, late menopause (Franceschi *et al*, 1991; Tsilidis *et al*, 2011a), early menarche (Gong *et al*, 2013), nulliparity (Adami *et al*, 1994; Tsilidis *et al*, 2011a), miscarriage (Braem *et al*, 2012), hormone replacement therapy (HRT) (Beral *et al*, 2007; Greiser *et al*, 2007; Tsilidis *et al*, 2011b), endometriosis (Pearce *et al*, 2012; Heidemann *et al*, 2014), genital powder use (Terry *et al*, 2013), family history of breast/ovarian cancer (Cramer *et al*, 1983; Kazerouni *et al*, 2006), body mass index (BMI; Olsen *et al*, 2013), cigarette smoking (Faber *et al*, 2013), and pre-existing diabetes (Lee *et al*, 2013) have been reported to increase the ovarian cancer risk, whereas breast-feeding (Jordan *et al*, 2012), contraception (oral contraceptives (OCs), intrauterine device (IUD), and tubal ligation) (Collaborative Group on Epidemiological Studies of Ovarian Cancer *et al*, 2008; Ness *et al*, 2011; Tsilidis *et al*, 2011a; Rice *et al*, 2012; Sieh *et al*, 2013; Rice *et al*, 2014), hysterectomy and unilateral ovariectomy (Rice *et al*, 2012, 2014), and possibly alcohol drinking (Genkinger *et al*, 2006; Kelemen *et al*, 2013) have shown protective effects. A risk prediction model based on a selection of these factors could help identify women who reach a minimal risk level to benefit from targeted prevention measures such as cancer screening or use of chemopreventive agents. Recently, Pfeiffer *et al* (2013) have developed an ovarian risk prediction model for US women. Their model included parity, HRT, OC use, and family history of breast/ovarian cancer, and showed a modest discriminatory power (concordance statistic = 0.59) in external validation.

In the present study, we aimed to build an ovarian cancer risk prediction model for women in Western Europe using data from the European Prospective Investigation into Cancer and Nutrition (EPIC), with particular interest in examining whether the discriminatory power could be improved by considering more epidemiological risk factors.

MATERIAL AND METHODS

The EPIC cohort. The EPIC study is a multicentre, population-based cohort study including >520 000 participants (367 903 women), recruited between 1992 and 2000 in 23 study centres across 10 European countries (Norway, Sweden, Denmark, the United Kingdom, the Netherlands, Germany, France, Spain, Italy, and Greece). The study rationale and design have been described in detail elsewhere (Riboli and Kaaks, 1997; Riboli *et al*, 2002). At recruitment, all study participants completed questionnaires on lifestyle factors (including smoking history and alcohol drinking) and medical history, and for women, menstrual and reproductive histories, and history of HRT (details in Supplementary Table S1). Menopausal status at the time of enrolment was determined using information on recent menstrual cycles, hysterectomy, ovariectomy, and current HRT. Women aged 46–55 years with no or incomplete data to determine menopausal status were classified as peri-menopausal or unknown menopausal status. Baseline anthropometric data, including height, weight, and body circumference data were measured directly in all study centres except France, Norway, and the Oxford centre in the United Kingdom, where the data were self-reported. All study participants provided written informed consent, and the local ethics review boards of the participating institutions gave approval for the studies.

Prospective ascertainment of disease outcome and vital status. Through the follow-up period (until end of 2009, varying by centre), new cancer cases were identified through record linkage with regional or national cancer registries (Norway, Denmark, the United Kingdom, the Netherlands, Spain, and Italy), or by a combination of active follow-up, linkages to health insurance records and complementary requests of records from pathology registries (Germany, France, and Greece). In the present study, ovarian cancer was defined as ovarian, fallopian tube, and primary peritoneal cancer (ICD-O-2 codes C56.9, C57.0, and C48, respectively). Data on vital status were obtained from death registries at the regional or national level.

Exclusions. We excluded women with the following characteristics: (1) history of cancer (except non-melanoma skin cancer) at recruitment ($n = 19\,707$); (2) bilateral ovariectomy ($n = 10\,500$); (3) never menstruated ($n = 61$); (4) incomplete follow-up data ($n = 2205$); or (5) did not return the baseline questionnaire ($n = 509$). We further excluded women whose baseline age was <45 years ($n = 69\,215$) because the incidence rate of ovarian cancer in this age group is extremely low (Quirk *et al*, 2002), and some reproductive factors (i.e., parity) are subject to change among younger women. Participants from Norway and Sweden were excluded from analyses, as most of the reproductive risk factors considered in the present study were not collected in these countries. After these exclusions, the final study population consisted of 202 206 women.

Statistical analysis. After reviewing the literature and considering available data, we identified the following factors as candidate predictors: menopausal status, age at menopause, age at menarche, number of full-term pregnancies (FTPs), age at first FTP, duration of breast-feeding, number of miscarriages, unilateral ovariectomy, hysterectomy, HRT, OC use, IUD use, BMI, smoking status, alcohol consumption, and pre-existing diabetes. Information on family history of breast cancer was missing for 53.0% of the study participants, and information on family history of ovarian cancer was not collected, thus these two factors were not considered in our model building process.

Multiple imputation of missing data. Missing data of individual candidate predictors were mostly sporadic and occurred in all participating study centres. However, a complete-data analysis would contain only 120 827 subjects (including 453 cases). To avoid this loss, we imputed the missing values with five-fold multivariate imputation by chained equations (Buuren and Groothuis-Oudshoorn, 2013), resulting in five complete data sets. All statistical analyses were conducted on the five data sets. The imputation quality is summarised in Supplementary Table S2.

Derivation of the risk prediction model. We built a cause-specific competing risk model from two components. As one part, the risk of incident ovarian cancer was estimated with a multivariable Weibull model. This was adjusted for the competing risk of death or other incident cancer with a Gompertz model as a second part in a modular fashion, as suggested by Benichou and Gail (1990).

All models were fit on each of the five imputed data sets, stratified by country. Age was used as the underlying timescale and observations were regarded as left-truncated at recruitment.

The development of incident ovarian cancer was estimated with a multivariable Weibull proportional hazards (PH) model (Collet, 2003) including all candidate predictors. The choice of this parametric model was based upon graphical comparisons with the non-parametric Nelson–Aalen estimate of the cumulative incidence (Supplementary Figure S1 in Supplementary Material). Age at exit was defined as the age at diagnosis of any cancer (except non-melanoma skin cancer), death from non-cancer causes,

withdrawal from the cohort, or end of follow-up, whichever came first. Age at menopause, duration of HRT, duration of OC use, number of FTPs (0, 1, 2, 3, and 4+), duration of breast-feeding, and BMI were included in the models as continuous variables. Power transformations of these variables (Royston *et al*, 1999) did not improve the model fit (Supplementary Table S3). We created indicators for menopausal status, parity, HRT use, and OC use, and centred age at menopause, number of FTPs, age at the first FTP, duration of HRT use, and duration of OC use. Indicators for menopausal status, parity, HRT use, and OC use were coded as interaction parameters to reflect the conditional relations among variables. Information regarding how the candidate predictors were coded is detailed in Supplementary Table S4.

We fit one model including all the described risk factors. To derive a parsimonious model of similar prediction quality, we also performed backward elimination within the Weibull PH model. To preserve all important predictors, we used an inclusion criterion of $P \leq 0.1$ and finally retained any predictors selected by at least three of the five imputed data sets (Vergouwe *et al*, 2010). Eventually, a Weibull PH model including the final predictors was fitted on each of the five imputed data sets, and the parameter estimates were combined according to Rubin's rules (Rubin, 1987). A relative risk score was calculated as the sum of the products of the individual predictor values and the associated linear covariate parameter estimates. We found no violation of the proportionality assumption for the final predictors from the Schoenfeld residuals (Supplementary Table S5).

Incident cancers other than primary ovarian cancer and deaths due to non-cancer causes were modelled as competing events with a country-stratified Gompertz model using age as the timescale. Again the parametric model class was decided upon after graphical comparison with the non-parametric Nelson–Aalen estimate, and estimates from the imputed data sets were combined under Rubin's rules (Supplementary Figure S2 in Supplementary Material). To estimate the absolute 5-year risk of developing ovarian cancer for women at age t in a cause-specific competing risk context, we finally used the following equation adapted from Benichou and Gail (1990):

$$F(t, t+5; x) = \int_t^{t+5} f(u) = \int_t^{t+5} h_1(u; x) \exp\left[-\int_t^u \{h_1(v; x) + h_2(v)\} dv\right] du$$

In this equation, $h_1(u; x)$ is the multivariable Weibull-hazard function for incident ovarian cancer, $h_2(v)$ is the Gompertz-hazard function for the competing events, and the term $\exp\left[-\int_t^u \{h_1(v; x) + h_2(v)\} dv\right]$ expresses the probability of remaining alive and free of any cancer at age u .

Assessment of model quality with cross-validation. We performed a five-fold cross-validation to obtain estimates of model predictive accuracy from data sections that were kept aside from model fitting. The discriminatory power was evaluated using the overall concordance index (C-index), a modification to the area under the receiver-operating characteristic curve (AUROC) adapted to survival data (Pencina and D'Agostino, 2004). To assess the overall calibration, we calculated the ratio of the expected (E) to the observed number (O) of incident ovarian cancer cases within the first 5 years of follow-up. The expected number of cases was calculated as the sum of the 5-year ovarian cancer absolute risk among women who either remained alive and free of any cancer or developed ovarian cancer or competing events in the first 5 years of follow-up. The 95% confidence interval (CI) of the E/O was calculated with $(E/O) \times \exp\left(\pm 1.96 \times \sqrt{\frac{1}{O}}\right)$. We also assessed the agreement between the estimated and observed numbers of cases across each tenth of the predicted 5-year absolute risk using the Hosmer-Lemeshow (H-L) test and provide the corresponding calibration plot. In addition, the calibration slope was calculated

from a regression of the cross-validation data, again combining estimates across multiple imputation with Rubin's rules.

Missing data imputation was performed using the 'mice' package in R (version 3.0.1, R Foundation for Statistical Computing, Vienna, Austria). The Weibull models and the Gompertz models were fitted with the R package 'eha'. The other statistical analyses were performed using SAS (version 9.2, SAS Institute, Cary, NC, USA).

RESULTS

After a median follow-up time of 11.7 years (range: 0.1–16.6), 791 primary ovarian cancers were diagnosed (median age at diagnosis: 63.4 years; range: 45.6–98.7), of them 324 cases were diagnosed within the first 5 years of follow-up. Other primary cancers were diagnosed in 9975 women, and 6386 women died due to non-cancer causes. Distribution of the baseline characteristics of the study population is presented in Table 1. Median age at recruitment was 52.4 years (range: 45.0–77.8). A total of 61.5% of women were postmenopausal, and the median age at menopause was 50 years (range: 12–67). Nearly 30% of women had ever used HRT and more than half (51.9%) of women had taken OCs. A relatively high frequency of missing values was observed for age at menopause (among postmenopausal women, 22.6%), duration of OC use (among ever users, 12.7%), and number of miscarriages (10.1%). Baseline BMI was on average 24.5 kg m^{-2} (range: 13.0–74.5).

Risk factor effects were little affected by model selection, as can be seen from the parameter estimates and hazard ratios (Table 2). After backward selection, the risk factors menopausal status and age at menopause, HRT use and duration of HRT, OC use and duration of OC use, parity and number of FTPs, unilateral ovariectomy, and BMI remained in the model. Older age at menopause, longer duration of HRT use, and higher BMI were associated with an increased ovarian cancer risk, whereas OC use, longer duration of OC use, parity, more FTPs, and unilateral ovariectomy were associated with a reduced risk. Using these predictors' coefficients, a woman's ovarian cancer relative risk score was calculated as: $RR = \exp(0.019 \times \text{meno_stat} + 0.034 \times \text{meno_age} + 0.086 \times \text{hrt} + 0.057 \times \text{hrt_dur} - 0.181 \times \text{oc} - 0.034 \times \text{oc_dur} - 0.308 \times \text{parity} - 0.094 \times \text{ftps} - 0.691 \times \text{uni_ovariect} + 0.021 \times \text{bmi})$.

The country-specific Weibull and Gompertz parameter estimates are provided as Supplementary Table S6. The deciles of the relative risk score are translated into absolute risk levels across an age range from 45 to 80 years (Figure 1). For women with the lowest and the highest tenth relative risk score, the corresponding 5-year risk at 45 years of age was 0.04% and 0.10%, respectively. The highest risk was observed at 68 years of age, ranging from 0.10% for the lowest tenth risk score to 0.24% for the highest tenth risk score, followed by a decline in older age groups.

The C-index from cross-validation was 0.64 (95% CI: 0.58, 0.70) for the full model and 0.64 (95% CI: 0.57, 0.70) for the selected model (Table 3), implying a modest discrimination. With respect to calibration, the selected model predicted 293 ovarian cancers to occur during the first 5 years of follow-up, in contrast to 324 cases that were actually observed ($E/O = 0.90$; 95% CI: 0.81, 1.01). This underestimation can be observed in eight decile groups in the calibration plot of the selected model (Figure 2), although the H-L test gave no evidence for miscalibration in general ($P = 0.14$). With estimates ~ 0.9 , the calibration slope confirmed the above described tendency of overfitting, but the CIs do not indicate significance.

We also externally validated Pfeiffer's model (Pfeiffer *et al*, 2013) in a subgroup of EPIC women ($n = 66\,493$) who had

Table 1. Distribution of baseline characteristics among the study population in the EPIC cohort presented as median (range) or percent

	All (n = 20 2206)	France (n = 59 981)	Italy (n = 21 727)	Spain (n = 14 153)	UK (n = 31 943)	Netherlands (n = 20 053)	Greece (n = 10 133)	Germany (n = 16 259)	Denmark (n = 27 957)
No. of cases	791	144	77	46	190	90	36	63	145
Age at entry (median, range)	52.4 (45.0, 77.8)	52.4 (45.0, 71.4)	53.9 (45.0, 77.8)	52.8 (45.0, 69.8)	56.1 (45.0, 98.5)	55.3 (45.0, 70.1)	60.1 (45.0, 84.4)	55.0 (45.0, 70.1)	56.2 (50.1, 65.8)
Menopause status (%)									
Post Pre/perio/ unknown	61.5 38.5	49.4 50.6	59.8 40.2	56.3 43.7	66.3 33.7	64.3 35.7	78.3 21.7	64.1 35.9	76.3 23.7
Age at menopause (median, range) ^a	50 (12, 67)	50 (17, 67)	50 (15, 62)	50 (25, 61)	50 (12, 67)	50 (16, 64)	49 (23, 66)	50 (45, 61)	50 (13, 64)
% Missing	22.6	45.7	6.2	4.3	27.5	9.6	4.5	19.2	18.9
HRT use (%)									
Never Ever Missing	66.9 29.8 3.3	66.1 33.2 0.7	79.3 19.1 1.6	85.9 11.7 2.4	64.4 31.2 4.4	75.2 23.2 1.6	94.1 5.3 0.6	33.1 47.0 19.9	55.8 42.4 1.8
Duration of HRT use (years) ^b	2.2 (0.1, 38)	2.2 (0.1, 30)	1.0 (0.5, 20)	1.0 (0.5, 20)	2.2 (0.1, 38)	2.0 (0.5, 30)	0.8 (0.1, 18)	4.0 (0.5, 20)	4.0 (0.5, 20)
% Missing	3.1	0.02	6.6	2.1	7.3	7.1	1.9	4.8	6.0
Use of OCs (%)									
Never Ever Missing	46.9 51.9 1.2	41.0 58.8 0.2	65.6 33.0 1.4	71.0 28.9 0.1	45.1 50.3 4.6	32.0 67.3 0.7	94.1 5.4 0.5	25.9 73.9 0.2	40.8 58.1 1.1
Duration of OC use (years) ^c (median, range)	5 (1, 25)	5 (1, 25)	2 (1, 25)	2 (1, 25)	5 (1, 25)	8 (1, 25)	1 (1, 25)	10 (1, 25)	6 (1, 25)
% Missing	12.7	27.4	3.2	1.0	11.4	3.9	3.7	0.7	5.6
FTP (%)									
0 ≥1 Missing	11.0 83.6 5.4	8.8 83.4 7.8	11.8 87.0 1.2	9.9 89.2 0.9	13.8 82.1 4.1	12.6 65.8 21.6	8.4 91.0 0.6	12.8 87.0 0.2	11.4 88.3 0.3
Age at first FTP (median, range) ^d % Missing	25 (12, 56) 0.4	24 (15, 48) 0.3	25 (13, 48) 0.1	25 (13, 48) 0.1	25 (12, 56) 0.4	25 (12, 45) 0.05	24 (14, 49) 1.2	23 (15, 44) 0.1	23 (14, 55) 1.1
Duration of breast- feeding (months) (median, range) ^d % Missing	4 (0, 286) 3.2	2.5(0,99) 5.8	6 (0, 84) 0.2	9(0, 144) 0.3	3.7 (0, 200) 6.5	2.8 (0, 107) 0.6	12 (0, 286) 1.2	2 (0, 84) 0.6	6 (0, 150) 2.1
Multiple miscarriages (%)									
No Yes Missing	79.7 5.2 10.1	79.9 4.9 15.2	83.6 5.5 10.9	82.1 7.8 10.1	76.4 4.6 19.0	61.5 4.7 33.8	85.4 8.0 6.6	58.0 4.8 10.2	86.5 4.9 8.6
Age at menarche (median, range)	13 (8, 20)	13 (8, 20)	13 (8, 20)	13 (8, 20)	13 (8, 20)	13 (8, 20)	13 (8, 20)	13 (8, 20)	14 (8, 20)
Use of intrauterine device (%)									
Never Ever Missing	70.8 25.3 3.9	60.0 39.9 0.1	84.3 14.3 1.4	90.4 8.7 0.9	72.9 22.1 5.0	65.3 9.2 25.5	97.3 2.1 0.6	70.1 27.1 2.8	65.7 33.6 0.7
Hysterectomy (%)									
No Yes Missing	87.0 11.2 1.8	88.7 8.2 3.1	91.7 6.8 1.5	95.6 4.3 0.1	80.4 15.5 4.1	79.5 19.9 0.6	95.6 3.9 0.5	81.8 18.2 0.0	88.4 11.4 0.2
Unilateral ovariectomy (%)									
No Yes Missing	92.1 4.7 3.2	91.4 3.6 5.0	93.6 4.7 1.7	96.6 3.2 0.2	89.0 4.5 6.5	90.0 7.0 3.0	97.0 1.9 1.1	93.2 6.7 0.1	92.9 6.0 1.1
BMI (median, range)	24.5 (13.0, 74.5)	22.6 (13.2, 57.9)	25.5 (15.0, 69.4)	28.6 (17.1, 62.2)	24.5 (13.4, 74.5)	24.9 (13.1, 52.8)	29.1 (15.4, 67.1)	25.6 (13.0, 58.7)	24.8 (13.4, 58.6)

Table 1. (Continued)

	All (n = 20 2206)	France (n = 59 981)	Italy (n = 21 727)	Spain (n = 14 153)	UK (n = 31 943)	Netherlands (n = 20 053)	Greece (n = 10 133)	Germany (n = 16 259)	Denmark (n = 27 957)
Smoking status (%)									
Never	60.1	67.2	56.6	85.0	55.7	41.4	83.2	62.2	43.9
Former	21.7	18.7	18.5	5.8	31.2	33.3	4.3	23.3	24.6
Current	15.6	8.5	23.6	9.2	9.7	24.7	9.1	14.3	31.3
Missing	2.6	5.6	1.3	0.0	3.4	0.6	3.4	0.2	0.2
Alcohol intake (g per day; median, range)	4 (0, 382)	6 (0, 382)	2 (0, 133)	0 (0, 115)	3 (0, 100)	4 (0, 164)	0.6 (0, 180)	5 (0, 263)	9 (0, 195)
% Missing	1.4	0	1.3	0	7.4	0.5	0	0	0.1
Family history of breast cancer (%)									
No	42.9	77.0	0	93.9	40.3	71.6	0	0	0
Yes	4.1	7.5	0	4.8	2.7	11.5	0	0	0
Missing	53.0	15.5	100	1.3	57.0	16.9	100	100	100
Pre-existing diabetes (%)									
No	94.9	97.8	96.2	93.7	89.5	96.7	90.4	95.9	94.4
Yes	2.8	2.1	2.4	6.1	1.6	2.6	8.9	4.0	1.5
Missing	2.3	0.1	1.4	0.2	8.9	0.7	0.7	0.1	4.1

Abbreviations: BMI = body mass index; EPIC = European Prospective Investigation into Cancer and Nutrition; FTP = full-term pregnancy; HRT = hormone replacement therapy; OC = oral contraceptive.
^aAmong postmenopausal women.
^bAmong ever HRT users.
^cAmong ever OC users.
^dAmong parous women.

Table 2. Combined HRs and 95% CIs and β -coefficients for risk predictors in the full and in the selected model

	Full model		Selected model	
	HR (95% CI)	Coefficient	HR (95% CI)	Coefficient
Age at menarche (1-year increase)	1.01 (0.92, 1.10)	0.008		
Parity (yes vs no)	0.78 (0.63, 0.96)	-0.248	0.74 (0.61, 0.89)	-0.308
No. FTPs (1-FTP increase)	0.93 (0.83, 1.04)	-0.075	0.91 (0.83, 1.00)	-0.094
Age at first live birth (1-year increase)	0.99 (0.98, 1.01)	-0.006		
Duration of breast-feeding (1-month increases)	0.99 (0.98, 1.00)	-0.008		
Miscarriages (≤ 1 vs ≥ 2)	1.29 (0.86, 1.94)	0.258		
OC use (yes vs no)	0.83 (0.70, 0.99)	-0.181	0.83 (0.70, 0.99)	-0.181
Duration of OC use (1-year increase)	0.97 (0.95, 0.98)	-0.035	0.97 (0.95, 0.98)	-0.034
Intrauterine device (never vs ever)	1.03 (0.85, 1.25)	0.033		
Menopausal status: post vs pre/peri	1.00 (0.81, 1.23)	0.001	1.02 (0.83, 1.25)	0.019
Age at menopause (1-year increase)	1.03 (1.01, 1.06)	0.033	1.03 (1.01, 1.06)	0.034
HRT use (yes vs no)	1.09 (0.92, 1.30)	0.088	1.09 (0.92, 1.29)	0.086
Duration of HRT use (1-year increase)	1.06 (1.03, 1.08)	0.057	1.06 (1.03, 1.08)	0.057
Hysterectomy (yes vs no)	0.90 (0.70, 1.16)	-0.106		
Unilateral ovariectomy (yes vs no)	0.51 (0.33, 0.79)	-0.670	0.50 (0.33, 0.77)	-0.691
BMI (1-unit increase)	1.02 (1.00, 1.04)	0.021	1.02 (1.00, 1.04)	0.021
Prevalent diabetes	1.09 (0.72, 1.64)	0.087		
Smoking status				
Former vs never	0.96 (0.80, 1.15)	-0.039		
Current vs never	1.16 (0.95, 1.42)	0.149		
Alcohol intake				
<6 vs 0g per day	1.10 (0.88, 1.37)	0.095		
6-11.9 vs 0g per day	0.93 (0.71, 1.22)	-0.072		
12-23.9 vs 0g per day	1.01 (0.76, 1.32)	0.006		
≥ 24 vs 0g per day	0.96 (0.70, 1.30)	-0.042		

Abbreviations: BMT = body mass index; CI = confidence interval; FTP = full-term pregnancy; HR = hazard ratio; HRT = hormone replacement therapy; OC = oral contraceptive.

information on family history of breast cancer, resulting in an overall C-index of 0.55 (95% CI: 0.52, 0.59). Our model performed on this subset with a C-index of 0.63 (CI: 0.51-0.76)

and E/O ratio of 0.90 (95%-CI: 0.74-1.08). The E/O for the Pfeiffer's model was 1.35 (95% CI: 1.12, 1.63), indicating significant overestimation.

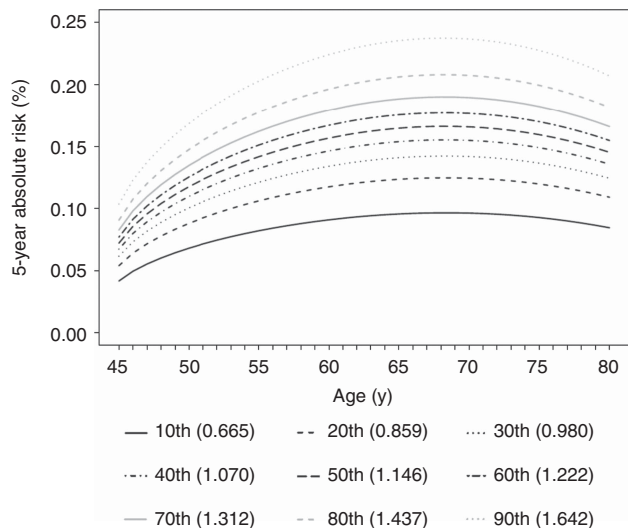


Figure 1. Predicted age-specific 5-year absolute ovarian cancer risk at decile cutoffs of the relative risk score. $RR = \exp(0.019 \times \text{meno_stat} + 0.034 \times \text{meno_age} + 0.086 \times \text{hrt} + 0.057 \times \text{hrt_dur} - 0.181 \times \text{oc} - 0.034 \times \text{oc_dur} - 0.308 \times \text{parity} - 0.094 \times \text{ftps} - 0.691 \times \text{uni_ovariect} + 0.021 \times \text{bmi})$. Country effect was fixed at the average level.

Table 3. Discrimination and calibration of full and selected model from five-fold cross-validation

	C-index (95% CI)	E/O ratio (95% CI)	Calibration slope (95% CI)
Full model	0.64 (0.58, 0.70)	0.91 (0.82; 1.01)	0.88 (0.64; 1.11)
Selected model	0.64 (0.57, 0.70)	0.90 (0.81; 1.01)	0.90 (0.66; 1.15)

Abbreviation: CI = confidence interval.

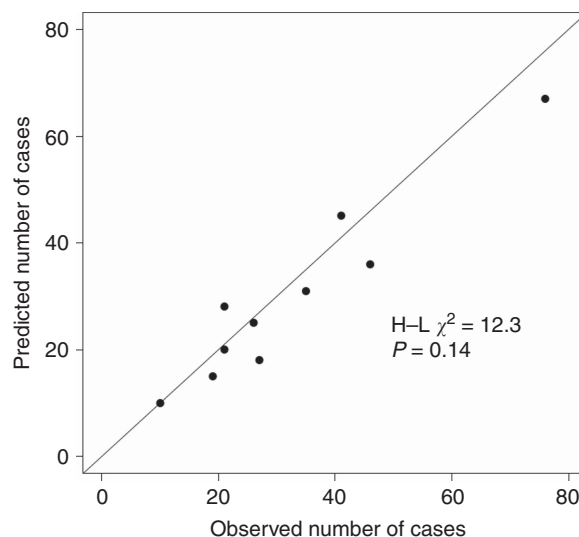


Figure 2. Predicted versus observed number of cases in the first 5 years of the follow-up by risk deciles.

DISCUSSION

We developed a risk prediction model for ovarian cancer based on epidemiological questionnaire data from European women aged 45 years and over. The risk factors menopausal status, age at menopause, duration of HRT, duration of OC use, unilateral ovariectomy, number of FTPs, and BMI were selected as major predictors. Cross-validation indicated that this model’s discriminatory power was modest (C-index: 0.64; 95% CI: 0.57, 0.70). Our model showed acceptable internal calibration, although the absolute risk was somewhat underestimated (E/O = 0.90; 95% CI: 0.81, 1.01).

In preventive oncology, an important use of ovarian cancer risk prediction models is to identify women who are at high risk and thus may benefit from targeted interventions, such as population-based screening programs (van Nagell and Hoff, 2013; Menon, Griffin, Gentry-Maharaj, 2014), or chemoprevention (e.g., use of low-dose, non-steroidal anti-inflammatory drugs (Baandrup *et al*, 2013; Trabert *et al*, 2014)). To our knowledge, only two epidemiologic risk models so far have been developed. The recent one was developed by Pfeiffer *et al* (Pfeiffer *et al*, 2013), incorporating the data that had been used to develop the earlier model by Rosner *et al* (2005). The predictors in the Pfeiffer model were also included in our model except for family history of breast/ovarian cancer. We observed that including additional predictors (namely, age at menopause, unilateral ovariectomy, and BMI) slightly improved the discriminatory power, compared with the modest discriminatory statistic of ~0.60 for the two US models. Application of the Pfeiffer model to the eligible part of our data showed less discriminatory ability than previously reported and substantial overestimation of the 5-year absolute risk, which may have been due to limited generalisability of predictive models between US and European populations. In the presence of the selected risk factors, other risk effects such as smoking, alcohol intake, or prevalent diabetes did not contribute independently to the predictive capacity of our model. The modest performance of the model is comparable to the discriminative potential of existing breast cancer risk-assessment tools, which also have a C-statistic

~ 60%. Such models are applied in prevention programs that have a low invasive character in general, including careful surveillance (as in chemoprevention (e.g., USPSTF, 2002)), and are also used as inclusion criteria in prevention trials (e.g., McCaskill-Stevens *et al*, 2013).

While the comparatively large study population and high number of ovarian cancer cases were the strengths of our model development, our study had several limitations that may compromise the predictive accuracy of our model. First of all, the EPIC cohort lacked or had incomplete information on several risk factors for ovarian cancer, first among these family history of breast/ovarian cancer, as ovarian cancer is strongly associated with mutations in BRCA1/2 genes (Powell, 2014). Tubal ligation, endometriosis, talcum use, and other possible risk factors such as polycystic ovary syndrome (Barry *et al*, 2014) were other factors lacking from our database. We also had no updated information on bilateral ovariectomy after recruitment, which may affect the competing risk estimates and model calibration but should not affect the discriminatory power. Regarding HRT, it has been suggested that oestrogen-only therapy has a stronger association with ovarian cancer risk than oestrogen/progestin therapy (Greiser *et al*, 2007; Tsilidis *et al*, 2011b), therefore distinguishing that HRT subtypes may improve the predictive accuracy. However, the information on use of specific HRT subtypes was only available for the current users at baseline. Finally, the number of ovarian cancer cases in our study was not large enough for more detailed risk modelling by ovarian cancer subtypes, as defined by histology or molecular (‘type-I/type-II’) characteristics (Shih and Kurman, 2004; see Supplementary Table S7 for details on subtypes). In a large US case-control study (1571 cases), a history of endometriosis and parity, having a previous tubal ligation and previous hysterectomy, showed a stronger association with endometrioid/

clear cell ovarian tumours or tumours classified as type-I (Merritt *et al*, 2013), implying that it might be more appropriate to build subtype-specific risk models for targeted prevention rather than a single model shared by all subtypes.

Our ovarian cancer risk prediction model has not been externally validated. However, we applied cross-validation to estimate model validity from parts of the data that were not used for model building, and this added credibility to our findings. Calibration towards country of origin was necessary in the model building process; this implies that the reported variation in ovarian cancer incidence rates in European countries (Ferlay *et al*, 2013) persisted even after adjustment for the risk factors considered in the model building process. Thus, model generalisation is difficult and validity may be impaired by population heterogeneity, as can also be seen in the poor performance of the Pfeiffer model on our data, although that model performed well on US cohort data.

An ovarian cancer risk prediction model based on clinical symptoms such as abdominal pain and rectal bleeding was first developed by Goff *et al* (2007) on a US-American case-control study showing promising levels of sensitivity (57% and 80% for early-stage and advanced-stage disease, respectively) and specificity (87% and 90% for women below and above the age of 50 years). A similar model recently developed in a UK study model showed a fairly high discriminatory accuracy (AUROC = 0.84) with regard to 2-year risk in the internal validation (Hippisley-Cox and Coupland, 2012), suggesting that the predictive accuracy of our model could be improved by adding (pre-) clinical symptoms. However, the expanded model may not perform well in early identification of women who are at high risk yet asymptomatic.

In summary, the ovarian cancer risk model we built using non-invasively measured epidemiologic risk factors showed a modest discriminatory power in a Western European population, comparable to previously developed models on US cohort data. Future studies should consider adding informative biomarkers to possibly improve the predictive ability of the model.

ACKNOWLEDGEMENTS

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, Mutuelle Générale de l'Éducation Nationale, Institut National de la Santé et de la Recherche Médicale (France); Deutsche Krebshilfe, Deutsches Krebsforschungszentrum and Federal Ministry of Education and Research (Germany); the Hellenic Health Foundation (Greece); Italian Association for Research on Cancer and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports, Netherlands Cancer Registry, LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund, Statistics Netherlands (The Netherlands); Norwegian Research Council, Norwegian Cancer Society, ERC-2009-AdG 232997 and Nordforsk, Nordic Centre of Excellence programme on Food, Nutrition and Health. (Norway); Health Research Fund (FIS), The Spanish Ministry of Health (ISCIII RETICC RD06/0020/0091), and the Catalan Institute of Oncology, Regional Governments of Andalucía, Asturias, Basque Country, Murcia (no. 6236), and Navarra, ISCIII RETIC (RD06/0020; Spain); Swedish Cancer Society, Swedish Scientific Council, and Regional Government of Skåne and Västerbotten (Sweden); Cancer Research UK, Medical Research Council (United Kingdom). KT and DC were supported by the NIH grant entitled 'Screening and Risk Biomarkers for Ovarian Cancer in EPIC Specimens' (1R01CA158119).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHORS CONTRIBUTIONS

KL and AH wrote the statistical analysis plan; KL performed the statistical analyses; KL, AH, and RK drafted the manuscript; RTF, AT, LH, LD, JC-C, MB, AS, CB, DT, AT, DP, AM, CA, RT, NCO-M, PHP, HB(as).B-d-M, ITG, EW, ES-C, M-DC, EJD, EA, AI, EL, K-TK, RCT, MAM, MJG, ER, PF, KT, DC, and RK conducted the epidemiologic studies and/or revised the paper. All authors critically reviewed the drafts of the manuscript and contributed to the writing.

DISCLAIMER

The sponsors had no role in study design, data collection, statistical analysis, interpretation of data, or writing of the manuscript.

REFERENCES

- Adami HO, Hsieh CC, Lambe M, Trichopoulos D, Leon D, Persson I, Ekblom A, Janson PO (1994) Parity, age at first childbirth, and risk of ovarian cancer. *Lancet* **344**(8932): 1250–1254.
- American Cancer Society (2014) <http://www.cancer.org/cancer/ovariancancer/detailedguide/ovarian-cancer-survival-rates>.
- Baandrup L, Faber MT, Christensen J, Jensen A, Andersen KK, Friis S, Kjaer SK (2013) Non-steroidal anti-inflammatory drugs and risk of ovarian cancer: systematic review and meta-analysis of observational studies. *Acta Obstet Gynecol Scand* **92**(3): 245–255.
- Barry JA, Azizia MM, Hardiman PJ (2014) Risk of endometrial, ovarian and breast cancer in women with polycystic ovary syndrome: a systematic review and meta-analysis. *Hum Reprod Update* **20**(5): 748–758.
- Benichou J, Gail MH (1990) Estimates of absolute cause-specific risk in cohort studies. *Biometrics* **46**(3): 813–826.
- Beral V. Million Women Study Collaborators Bull D, Green J, Reeves G (2007) Ovarian cancer and hormone replacement therapy in the Million Women Study. *Lancet* **369**(9574): 1703–1710.
- Braem MG, Onland-Moret NC, Schouten LJ, Kruitwagen RF, Lukanova A, Allen NE, Wark PA, Tjønneland A, Hansen L, Braüner CM, Overvad K, Clavel-Chapelon F, Chabbert-Buffet N, Teucher B, Floegel A, Boeing H, Trichopoulou A, Adarakis G, Plada M, Rinaldi S, Fedirko V, Romieu I, Pala V, Galasso R, Sacerdote C, Palli D, Tumino R, Bueno-de-Mesquita HB, Gram IT, Gavriluyuk O, Lund E, Sánchez MJ, Bonet C, Chirlaque MD, Larrañaga N, Gurrea AB, Quirós JR, Idahl A, Ohlson N, Lundin E, Jirström K, Butt S, Tsilidis KK, Khaw KT, Wareham N, Riboli E, Kaaks R, Peeters PH (2012) Multiple miscarriages are associated with the risk of ovarian cancer: results from the European Prospective Investigation into Cancer and Nutrition. *PLoS One* **7**(5): e37141.
- Buuren S, Groothuis-Oudshoorn K (2013) MICE: multivariate imputation by chained equations in R. *J Stat Softw* **45**.
- Collaborative Group on Epidemiological Studies of Ovarian Cancer Beral V, Doll R, Hermon C, Peto R, Reeves G (2008) 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* **371**(9609): 303–314.
- Collett D (2003) *Modelling survival data in medical research*. 2nd edn (Chapman & Hall/CRC: Florida, USA).
- Cramer DW, Hutchison GB, Welch WR, Scully RE, Ryan KJ (1983) Determinants of ovarian cancer risk. I. Reproductive experiences and family history. *J Natl Cancer Inst* **71**(4): 711–716.
- Faber MT, Kjaer SK, Dehendorff C, Chang-Claude J, Andersen KK, Høgdall E, Webb PM, Jordan SJ. Australian Cancer Study (Ovarian Cancer) Australian Ovarian Cancer Study Group Rossing MA, Doherty JA, Lurie G, Thompson PJ, Carney ME, Goodman MT, Ness RB, Modugno F, Edwards RP, Bunker CH, Goode EL, Fridley BL, Vierkant RA, Larson MC, Schildkraut J, Cramer DW, Terry KL, Vitonis AF, Bandera EV, Olson SH,

- King M, Chandran U, Kiemeny LA, Massuger LF, van Altena AM, Vermeulen SH, Brinton L, Wentzensen N, Lissowska J, Yang HP, Moysich KB, Odunsi K, Kasza K, Odunsi-Akanji O, Song H, Pharaoh P, Shah M, Whittemore AS, McGuire V, Sieh W, Sutphen R, Menon U, Gayther SA, Ramus SJ, Gentry-Maharaj A, Pearce CL, Wu AH, Pike MC, Risch HA, Jensen A. Ovarian Cancer Association Consortium (2013) Cigarette smoking and risk of ovarian cancer: a pooled analysis of 21 case-control studies. *Cancer Causes Control* **24**(5): 989–1004.
- Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, Rosso S, Coebergh JW, Comber H, Forman D, Bray F (2013) Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur J Cancer* **49**(6): 1374–1403.
- Franceschi S, La Vecchia C, Booth M, Tzonou A, Negri E, Parazzini F, Trichopoulos D, Beral V (1991) Pooled analysis of 3 European case-control studies of ovarian cancer: II. Age at menarche and at menopause. *Int J Cancer* **49**(1): 57–60.
- Genkinger JM, Hunter DJ, Spiegelman D, Anderson KE, Buring JE, Freudenheim JL, Goldbohm RA, Harnack L, Hankinson SE, Larsson SC, Leitzmann M, McCullough ML, Marshall J, Miller AB, Rodriguez C, Rohan TE, Schatzkin A, Schouten LJ, Wolk A, Zhang SM, Smith-Warner SA (2006) Alcohol intake and ovarian cancer risk: a pooled analysis of 10 cohort studies. *Br J Cancer* **94**(5): 757–762.
- Goff BA, Mandel LS, Drescher CW, Urban N, Gough S, Schurman KM, Patras J, Mahony BS, Andersen MR (2007) Development of an ovarian cancer symptom index: possibilities for earlier detection. *Cancer* **109**: 221–227.
- Gong TT, Wu QJ, Vogtmann E, Lin B, Wang YL (2013) Age at menarche and risk of ovarian cancer: a meta-analysis of epidemiological studies. *Int J Cancer* **132**(12): 2894–2900.
- Greiser CM, Greiser EM, Doren M (2007) Menopausal hormone therapy and risk of ovarian cancer: systematic review and meta-analysis. *Hum Reprod Update* **13**(5): 453–463.
- Heidemann LN, Hartwell D, Heidemann CH, Jochumsen KM (2014) The relation between endometriosis and ovarian cancer—a review. *Acta Obstet Gynecol Scand* **93**(1): 20–31.
- Hippisley-Cox J, Coupland C (2012) Identifying women with suspected ovarian cancer in primary care: derivation and validation of algorithm. *BMJ* **344**: d8009.
- Jordan SJ, Cushing-Haugen KL, Wicklund KG, Doherty JA, Rossing MA (2012) Breast-feeding and risk of epithelial ovarian cancer. *Cancer Causes Control* **23**(6): 919–927.
- Kazerouni N, Greene MH, Lacey Jr JV, Mink PJ, Schairer C (2006) Family history of breast cancer as a risk factor for ovarian cancer in a prospective study. *Cancer* **107**(5): 1075–1083.
- Kelemen LE, Bandera EV, Terry KL, Rossing MA, Brinton LA, Doherty JA, Ness RB, Kjaer SK, Chang-Claude J, Köbel M, Lurie G, Thompson PJ, Carney ME, Moysich K, Edwards R, Bunker C, Jensen A, Høgdall E, Cramer DW, Vitonis AF, Olson SH, King M, Chandran U, Lissowska J, Garcia-Closas M, Yang H, Webb PM, Schildkraut JM, Goodman MT, Risch HA. Australian Ovarian Cancer Study Group and Australian Cancer Study (Ovarian Cancer)Ovarian Cancer Association Consortium (2013) Recent alcohol consumption and risk of incident ovarian carcinoma: a pooled analysis of 5,342 cases and 10,358 controls from the Ovarian Cancer Association Consortium. *BMC Cancer* **13**: 28.
- Lee JY, Jeon I, Kim JW, Song YS, Yoon JM, Park SM (2013) Diabetes mellitus and ovarian cancer risk: a systematic review and meta-analysis of observational studies. *Int J Gynecol Cancer* **23**(3): 402–412.
- McCaskill-Stevens W, Wilson JW, Cook ED, Edwards CL, Gibson RV, McElwain DL, Figueroa-Moseley CD, Paskett ED, Roberson NL, Wickerham DL, Wolmark N (2013) National surgical adjuvant breast and bowel project study of tamoxifen and raloxifene trial: advancing the science of recruitment and breast cancer risk assessment in minority communities. *Clin Trials* **10**: 280–291.
- Menon U, Griffin M, Gentry-Maharaj A (2014) Ovarian cancer screening—current status, future directions. *Gynecol Oncol* **132**(2): 490–495.
- Merritt MA, De Pari M, Vitonis AF, Titus LJ, Cramer DW, Terry KL (2013) Reproductive characteristics in relation to ovarian cancer risk by histologic pathways. *Hum Reprod* **28**(5): 1406–1417.
- Ness RB, Dodge RC, Edwards RP, Baker JA, Moysich KB (2011) Contraception methods, beyond oral contraceptives and tubal ligation, and risk of ovarian cancer. *Ann Epidemiol* **21**(3): 188–196.
- Oberaigner W, Minicozzi P, Bielska-Lasota M, Allemanni C, de Angelis R, Mangone L, Sant M. Eurocare Working Group (2012) Survival for ovarian cancer in Europe: the across-country variation did not shrink in the past decade. *Acta Oncol* **51**(4): 441–453.
- Olsen CM, Nagle CM, Whiteman DC, Ness R, Pearce CL, Pike MC, Rossing MA, Terry KL, Wu AH, the Australian Cancer Study (Ovarian Cancer), Australian Ovarian Cancer Study Group, Risch HA, Yu H, Doherty JA, Chang-Claude J, Hein R, Nickels S, Wang-Gohrke S, Goodman MT, Carney ME, Matsuno RK, Lurie G, Moysich K, Kjaer SK, Jensen A, Høgdall E, Goode EL, Fridley BL, Vierkant RA, Larson MC, Schildkraut J, Hoyo C, Moorman P, Weber RP, Cramer DW, Vitonis AF, Bandera EV, Olson SH, Rodriguez-Rodriguez L, King M, Brinton LA, Yang H, Garcia-Closas M, Lissowska J, Anton-Culver H, Ziogas A, Gayther SA, Ramus SJ, Menon U, Gentry-Maharaj A, Webb PM. Ovarian Cancer Association Consortium (2013) Obesity and risk of ovarian cancer types: evidence from the Ovarian Cancer Association Consortium. *Endocr Relat Cancer* **20**(2): 251–262.
- Pearce CL, Templeman C, Rossing MA, Lee A, Near AM, Webb PM, Nagle CM, Doherty JA, Cushing-Haugen KL, Wicklund KG, Chang-Claude J, Hein R, Lurie G, Wilkens LR, Carney ME, Goodman MT, Moysich K, Kjaer SK, Høgdall E, Jensen A, Goode EL, Fridley BL, Larson MC, Schildkraut JM, Palmieri RT, Cramer DW, Terry KL, Vitonis AF, Titus LJ, Ziogas A, Brewster W, Anton-Culver H, Gentry-Maharaj A, Ramus SJ, Anderson AR, Brueggmann D, Fasching PA, Gayther SA, Huntsman DG, Menon U, Ness RB, Pike MC, Risch H, Wu AH, Berchuck A. Ovarian Cancer Association Consortium (2012) Association between endometriosis and risk of histological subtypes of ovarian cancer: a pooled analysis of case-control studies. *Lancet Oncol* **13**(4): 385–394.
- Pencina MJ, D'Agostino RB (2004) Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Statist Med* **23**(13): 2109–2123.
- Pfeiffer RM, Park Y, Kreimer AR, Lacey Jr JV, Pee D, Greenlee RT, Buys SS, Hollenbeck A, Rosner B, Gail MH, Hartge P (2013) 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* **10**(7): e1001492.
- Powell CB (2014) Clinical management of patients at inherited risk for gynecologic cancer. *Curr Opin Obstet Gynecol* **27**(1): 14–22.
- Quirk JT, Natarajan N, Mettlin CJ (2002) Age-specific ovarian cancer incidence rate patterns in the United States. *Gynecol Oncol* **99**(1): 248–250.
- Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, Fahey M, Charrondière UR, Hémond B, Casagrande C, Vignat J, Overvad K, Tjønneland A, Clavel-Chapelon F, Thiébaud A, Wahrendorf J, Boeing H, Trichopoulos D, Trichopoulou A, Vineis P, Palli D, Bueno-De-Mesquita HB, Peeters PH, Lund E, Engeset D, González CA, Barricarte A, Berglund G, Hallmans G, Day NE, Key TJ, Kaaks R, Saracci R (2002) European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr* **5**(6B): 1113–1124.
- Riboli E, Kaaks R (1997) The EPIC Project: rationale and study design. *European Prospective Investigation into Cancer and Nutrition. Int J Epidemiol* **26**(Suppl 1): S6–14.
- Rice MS, Hankinson SE, Tworoger SS (2014) Tubal ligation, hysterectomy, unilateral oophorectomy, and risk of ovarian cancer in the Nurses' Health Studies. *Fertil Steril* **102**(1): 192–198.
- Rice MS, Murphy MA, Tworoger SS (2012) Tubal ligation, hysterectomy and ovarian cancer: a meta-analysis. *J Ovarian Res* **5**(1): 13.
- Rosner BA, Colditz GA, Webb PM, Hankinson SE (2005) Mathematical models of ovarian cancer incidence. *Epidemiology* **16**(4): 508–515.
- Royston P, Ambler G, Sauerbrei W (1999) The use of fractional polynomials to model continuous risk variables in epidemiology. *Int J Epidemiol* **28**(5): 964–974.
- Rubin DB (1987) *Multiple Imputation for Nonresponse in Surveys*. Wiley: New York, USA.
- Shih L, Kurman R (2004) Ovarian tumorigenesis: a proposed model based on morphological and molecular genetic analysis. *Am J Pathol* **164**(5): 1511–1518.
- Sieh W, Salvador S, McGuire V, Weber RP, Terry KL, Rossing MA, Risch H, Wu AH, Webb PM, Moysich K, Doherty JA, Felberg A, Miller D, Jordan SJ. Australian Cancer Study (Ovarian Cancer)Australian Ovarian Cancer Study GroupGoodman MT, Lurie G, Chang-Claude J, Rudolph A, Kjaer SK, Jensen A, Høgdall E, Bandera EV, Olson SH, King MG, Rodriguez-Rodriguez L, Kiemeny LA, Marees T, Massuger LF, van Altena AM, Ness RB, Cramer DW, Pike MC, Pearce CL, Berchuck A, Schildkraut JM, Whittemore AS. Ovarian Cancer Association Consortium (2013) Tubal ligation and risk of ovarian cancer

- subtypes: a pooled analysis of case-control studies. *Int J Epidemiol* **42**(2): 579–589.
- Terry KL, Karageorgi S, Shvetsov YB, Merritt MA, Lurie G, Thompson PJ, Carney ME, Weber RP, Akushevich L, Lo-Ciganic WH, Cushing-Haugen K, Sieh W, Moysich K, Doherty JA, Nagle CM, Berchuck A, Pearce CL, Pike M, Ness RB, Webb PM. Australian Cancer Study (Ovarian Cancer) Australian Ovarian Cancer Study Group Rossing MA, Schildkraut J, Risch H, Goodman MT. Ovarian Cancer Association Consortium (2013) Genital powder use and risk of ovarian cancer: a pooled analysis of 8,525 cases and 9,859 controls. *Cancer Prev Res* **6**(8): 811–821.
- Trabert B, Ness RB, Lo-Ciganic WH, Murphy MA, Goode EL, Poole EM, Brinton LA, Webb PM, Nagle CM, Jordan SJ. Australian Ovarian Cancer Study Group Australian Cancer Study (Ovarian Cancer) Risch HA, Rossing MA, Doherty JA, Goodman MT, Lurie G, Kjær SK, Hogdall E, Jensen A, Cramer DW, Terry KL, Vitonis A, Bandera EV, Olson S, King MG, Chandran U, Anton-Culver H, Ziogas A, Menon U, Gayther SA, Ramus SJ, Gentry-Maharaj A, Wu AH, Pearce CL, Pike MC, Berchuck A, Schildkraut JM, Wentzensen N. Ovarian Cancer Association Consortium (2014) 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* **106**(2): djt431.
- Tsilidis KK, Allen NE, Key TJ, Dossus L, Kaaks R, Bakken K, Lund E, Fournier A, Dahm CC, Overvad K, Hansen L, Tjønneland A, Rinaldi S, Romieu I, Boutron-Ruault MC, Clavel-Chapelon F, Lukanova A, Boeing H, Schütze M, Benetou V, Palli D, Berrino F, Galasso R, Tumino R, Sacerdote C, Bueno-de-Mesquita HB, van Duijnhoven FJ, Braem MG, Onland-Moret NC, Gram IT, Rodríguez L, Duell EJ, Sánchez MJ, Huerta JM, Ardanaz E, Amiano P, Khaw KT, Wareham N, Riboli E (2011b) Menopausal hormone therapy and risk of ovarian cancer in the European prospective investigation into cancer and nutrition. *Cancer Causes Control* **22**(8): 1075–1084.
- Tsilidis KK, Allen NE, Key TJ, Dossus L, Lukanova A, Bakken K, Lund E, Fournier A, Overvad K, Hansen L, Tjønneland A, Fedirko V, Rinaldi S, Romieu I, Clavel-Chapelon F, Engel P, Kaaks R, Schütze M, Steffen A, Bamia C, Trichopoulou A, Zylis D, Masala G, Pala V, Galasso R, Tumino R, Sacerdote C, Bueno-de-Mesquita HB, van Duijnhoven FJ, Braem MG, Onland-Moret NC, Gram IT, Rodríguez L, Travier N, Sánchez MJ, Huerta JM, Ardanaz E, Larrañaga N, Jirström K, Manjer J, Idahl A, Ohlson N, Khaw KT, Wareham N, Mouw T, Norat T, Riboli E (2011a) Oral contraceptive use and reproductive factors and risk of ovarian cancer in the European Prospective Investigation into Cancer and Nutrition. *Br J Cancer* **105**(9): 1436–1442.
- USPSTF (2002) *Chemoprevention of breast cancer: recommendations and rationale* (online). Available at <http://www.uspreventiveservicestaskforce.org/3rduspstf/breastchemo/breastchemorr.htm>.
- van Nagell Jr JR, Hoff JT (2013) Transvaginal ultrasonography in ovarian cancer screening: current perspectives. *Int J Womens Health* **6**: 25–33.
- Vergouwe Y, Royston P, Moons KG, Altman DG (2010) Development and validation of a prediction model with missing predictor data: a practical approach. *J Clin Epidemiol* **63**: 205–214.
- This work is published under the standard license to publish agreement. After 12 months the work will become freely available and the license terms will switch to a Creative Commons Attribution-NonCommercial-Share Alike 4.0 Unported License.

Supplementary Information accompanies this paper on British Journal of Cancer website (<http://www.nature.com/bjc>)