An epidemiological model for prediction of endometrial cancer risk in Europe

A. Hüsing (1)*, L. Dossus (2,3,4), P. Ferrari (5), A. Tjønneland (6), L. Hansen (6), G. Fagherazzi (2,3,4), L. Baglietto (7,8), H. Schock (1), J. Chang-Claude (9), H. Boeing (10), A. Steffen (10), A. Trichopoulou (11,12), C. Bamia (13), M. Katsoulis (14), V. Krogh (15), D. Palli (16), S. Panico (17), N.C. Onland-Moret (18, 19), P. H. Peeters (18, 19), H. B. Bueno-de-Mesquita, (20, 21, 19), E. Weiderpass (22, 23, 24, 25), I. T. Gram (22), E. Ardanaz (26, 27), M. Obón-Santacana (28), C. Navarro (29, 27, 30), E. Sánchez-Cantalejo (31, 27), N. Etxezarreta (27, 32), N. E. Allen (33), K. T. Khaw (34), N. Wareham (35), S. Rinaldi (5), I. Romieu (5), M. A. Merritt (19), M. Gunter (19), E. Riboli (19), R. Kaaks (1)

1. Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany

2. Inserm, Centre for research in Epidemiology and Population Health (CESP), U1018, Nutrition, Hormones and Women's Health team, F-94805, Villejuif, France

3. UnivParis Sud, UMRS 1018, F-94805, Villejuif, France

4. IGR, F-94805, Villejuif, France

5. International Agency for Research on Cancer, Lyon, France.

6. Danish Cancer Society Research Center, Copenhagen, Denmark

7. Cancer Epidemiology Centre, Cancer Council Victoria, 3053, Melbourne, Australia

8. Centre for Molecular, Environmental, Genetic, and Analytic Epidemiology, The University of Melbourne, 3010, Melbourne, Australia

9. Unit of Genetic Epidemiology, Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany

10. German Institute of Human Nutrition in Potsdam-Rehbruecke, Potsdam, Germany

11. Hellenic Health Foundation, 13 Kaisareias Street, Athens, GR-115 27, Greece

12. Bureau of Epidemiologic Research, Academy of Athens, 23 Alexandroupoleos Street, Athens, GR-115 27, Greece

13. Department of Hygiene, Epidemiology and Medical Statistics, University of Athens Medical School, 75 M. Asias Street, Goudi, GR-115 27, Athens, Greece

14. Hellenic Health Foundation, 13 Kaisareias Street, Athens, GR-115 27, Greece

15. Epidemiology and Prevention Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Via Venezian, 1, 20133 Milano-Italy

16. Molecular and Nutritional Epidemiology Unit, Cancer Research and Prevention Institute – ISPO, Florence- Italy

17. Dipartimento Di Medicina Clinica E Chirurgia, Federico II University, Naples, Italy
18. Department of Epidemiology, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, The Netherlands
19. Department of Epidemiology and Biostatistics, School of Public Health, Imperial College, London, UK
20. National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands
21. Department of Gastroenterology and Hepatology, University Medical Centre, Utrecht, The Netherlands
22. Department of Community Medicine, Faculty of Health Sciences, University of Tromso, The Arctic University of Norway, Tromsø, Norway.
23. Department of Research, Cancer Registry of Norway, Oslo, Norway
24. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
25. Samfundet Folkhälsan, Helsinki, Finland
26. Navarre Public Health Institute, Pamplona, Spain.
27. CIBER Epidemiology and Public Health CIBERESP, Spain
28. Unit of Nutrition, Environment and Cancer. Cancer Epidemiology Research Program, Catalan Institute of Oncology (ICO-IDIBELL), Barcelona, Spain
29. Department of Epidemiology, Murcia Regional Health Council, Murcia, Spain
30. Department of Health and Social Sciences, Universidad de Murcia, Spain
31. Escuela Andaluza de Salud Pública. Instituto de Investigación Biosanitaria de Granada (Granada.ibs), Granada (Spain)
32. Public Health Division of Gipuzkoa-BIODONOSTIA, Basque Regional Health Department, Spain
33. CTSU, Nuffield Department of Population Health, University of Oxford, Oxford, UK.
34. University of Cambridge, School of Clinical Medicine, Cambridge, UK.
35. MRC Epidemiology Unit, University of Cambridge, Cambridge, UK.

Corresponding author:

Ms Anika Hüsing
German Cancer Research Center (DKFZ), Division of Cancer Epidemiology (C020)
Im Neuenheimer Feld 581, D-69120 Heidelberg
Phone: +49-(0)6221-42-2214,
Fax: +49-(0)6221-42-2203
E-Mail: a.huesing@dkfz.de

Abstract

Endometrial cancer is the fourth most frequent cancer in women in Europe, and as its incidence is increasing, prevention strategies gain further pertinence. Risk prediction models can be a useful tool for identifying women likely to benefit from targeted prevention measures. On the basis of data from 201,811 women (mostly aged 30-65 years) including 855 incident endometrial cancer cases from eight countries in the European Prospective Investigation into Cancer and Nutrition cohort, a model to predict endometrial cancer (EC) was developed. A step-wise model selection process was used to select confirmed predictive epidemiologic risk factors. Piece-wise constant hazard rates in 5-year age-intervals were estimated in a cause-specific competing risks model, five-fold-cross-validation was applied for internal validation.

Risk factors included in the risk prediction model were body-mass index (BMI), menopausal status, age at menarche and at menopause, oral contraceptive use, overall and by different BMI categories and overall duration of use, parity, age at first full-term pregnancy, duration of menopausal hormone therapy and smoking status (specific for pre, peri- and post-menopausal women). These variables improved the discriminating capacity to predict risk over 5 years from 71% for a model based on age alone to 77% (overall C statistic), and the model was well-calibrated (ratio of expected to observed cases = 0.99).

Our model could be used for the identification of women at increased risk of EC in Western Europe. To achieve an EC-risk model with general validity, a large-scale cohort-consortium approach would be needed to assess and adjust for population variation.

Introduction

Endometrial cancer (EC) is the fourth most frequent cancer in women in Europe (1). Most cases are diagnosed in older women, often after sixty years of age. Although EC is frequently diagnosed at an early stage when the disease is curable, about one fifth of all cases have a poor prognosis (1). Well-established risk factors for EC include obesity, use of estrogen-only hormone replacement therapy (HRT) and polycystic ovary syndrome, whereas the use of combined estrogen-plus-progestin HRT and oral contraceptives (OC) are protective (2-8). As the number of women living to an older age is steadily growing, together with the prevalence of obesity and sedentary lifestyles, the number of EC cases is also expected to increase in the near future (9). In view of these developments, prevention strategies gain further pertinence.

Information on a woman's risk of developing EC within the next five years derived from valid risk models can help physicians to motivate or proactively propose targeted clinical prevention measures such as the preventive use of progestin-based hormonal treatments (10). Risk models may help balance expected benefits against possible harms of preventive treatments and can also be a useful tool for selecting women who may benefit from screening programs (11, 12).

Recent investigations based on the prospective EPIC cohort have confirmed associations for several risk factors with EC, including reproductive factors (5), use of estrogen-alone *vs*. estrogen-plus-progestin HRT (4), excess body weight (13), physical inactivity (14), and cigarette smoking, which may have different effects in pre- and post-menopausal women (15). Taken together, these findings from the EPIC cohort led to the development of a risk prediction model for EC, combining the predictive capability from all of these risk factors. The availability of full prospective cohort data allowed direct estimation of absolute risk levels, and the large cohort size and number of incident cases of EC allowed application of robust methods for model building and assessment of discriminative capacity and calibration of absolute risk estimates.

Methods

The EPIC cohort

Study participants and baseline data collection

The European Prospective Investigation into Cancer and Nutrition (EPIC) is a multicenter, population-based cohort study including more than 520,000 participants (367,903 women), enrolled into the cohort between 1992 and 2000 in 23 study centers across 10 different countries (i.e. Denmark, France, Germany, Greece, Italy, the Netherlands, Norway, Spain, Sweden and

United Kingdom). The cohort population and data collection procedures have been described in detail elsewhere (16). All participating women completed questionnaires on their diet, lifestyle, menstrual and reproductive histories, and medical history. All study participants provided written informed consent, and the local ethical review boards of the participating institutions gave approval for the studies.

In most countries, height and weight were measured at recruitment according to standardized protocols. However, in EPIC-Oxford and approximately two-thirds of the French cohort, height and weight were self-reported. Body mass index was calculated as weight (kg) / (height (m))².

Data from the recruitment questionnaire included age at menarche, age at first and number of full-term pregnancies, duration of OC use, and for postmenopausal women additional questions addressed age at menopause and past or current HRT use. Women were considered premenopausal at baseline, if they reported having had regular menses over the past 12 months or if they were less than 46 years of age. Women who reported a bilateral ovariectomy or not having had any menses over the last 12 months or who were over 55 years old at recruitment were considered postmenopausal. Women between 46 and 55 years of age at recruitment with missing or incomplete questionnaire data on menopausal status were classified as peri-menopausal/unknown. Information on menopausal hormone-use at recruitment was used as indicator of ever use and cumulative duration.

Smoking behavior at baseline was summarized into categories of never, former, current or unknown. School level was considered a proxy for socioeconomic status. Information on self-reported diabetes, hypertension and hyperlipidemia at baseline was recorded in categories of yes, no and unknown.

Prospective ascertainment of cancer endpoints and vital status.

Incident cancer cases were identified through several methods, including record linkage with regional cancer registries (Denmark, Italy, the Netherlands, Spain and the United Kingdom), linkage to health insurance records, active follow-up of study subjects and systematic requests of patient records from pathology registries (France, Germany and Greece) (16). Data on vital status were obtained from mortality registries at the regional or national level.

Statistical Analyses

Exclusions

Participants who had reported hysterectomy (N=39,953), a prevalent cancer (except non-melanoma skin cancer) at the time of recruitment (N=15,317), or with missing follow-up data (lost to follow-up immediately after recruitment; N=2,591), were not considered in the analyses. Data from Norway and Sweden (N=61,119) were excluded, because several risk variables of interest had not been collected. Information on incident hysterectomy during follow-up was not available for this analysis. Simple imputation of sporadic missing values allowed us to use 201,811 participants with complete covariate data (81% of all eligible subjects).

Statistical Model

To derive an age specific absolute risk score from a number of risk factors we first developed a relative risk score. This relative risk score was used to calculate country- and age-specific baseline risk of EC. In a competing risk model these estimates were then combined with estimates for risk of hysterectomy and of developing other forms of primary cancer or death during follow-up, as detailed below.

Factors previously associated with EC risk in EPIC (4, 5, 13-15) were modeled in a Cox proportional hazards model with age as underlying time scale, starting observation at recruitment and ending at EC diagnosis or censored at end of follow-up or other event (death, other cancer). This analysis was stratified by center and age at recruitment in one-year increments. All factors previously associated with EC risk and all relevant interactions were combined into this model, together with potential confounders: menopausal status, BMI (continuous), interaction of BMI with menopausal status, age at first full term pregnancy (FTP) (centered at median age 24, nulliparous women coded 0), number of FTP in categories nulliparous, one, two, three or more, age at menarche (centered at median age 13), OC use, duration of OC use, interaction of OC use with BMI (for interaction regarded in categories normal: BMI<25, overweight: 25<=BMI <30 and obese: BMI>30), age at menopause (centered at median age 50, pre-/peri-menopausal women coded as 0), HRT use, duration of HRT use, interaction of HRT use with BMI (categorical), smoking status and interaction of smoking status with menopausal status, alcohol consumption with interaction with menopausal status, diabetes, hypertension and hyperlipidemia, physical activity and highest educational level.

A backwards elimination procedure was implemented, with inclusion criterion set to p=0.1. Linear shrinkage based on 100 bootstrap samples was applied to limit over-fitting. The resulting

"shrunk" parameter estimates were combined with the individuals' observed values into a one-dimensional relative risk score.

For the estimation of absolute risks of developing EC, the relative risk score was included into the estimation of a piece-wise-constant exponential baseline hazard function (including calibration parameters for country of origin), using 5-year age intervals as the underlying time-scale. Consistently with the Cox model described above, women who died or had another first primary incident cancer were censored at their age of that event. From the women's baseline data on prevalent hysterectomy and the age at which hysterectomy had been performed, we also estimated age- and country-specific risks of having a hysterectomy. These risk estimates were combined into a cause-specific competing risk model (17) for EC risk (probability $\pi(a, b, x)$, i.e. cumulative incidence to develop EC between age a and b with individual risk factor combination x) as summarized in formula (1):

$$\pi(a, b, x) = F(a, b, x) = \int_a^b \lambda_1(x) \cdot \exp(s \cdot (-\lambda_1(x) - \lambda_2 - \lambda_3)) \, ds \qquad (1)$$

In this equation $\lambda_1(x) = \lambda_{1\_a} \cdot \beta_{1,country} \cdot rr(x)$ represents the cause specific risk of EC calculated as product of (piecewise constant) baseline hazard λ_{1_a} in age-interval a, the constant factor $\beta_{1,country}$ to account for differences in incidence rates between countries, and the relative risk score derived from the Cox-model as $rr(x) = \exp(\beta_{rr}'x)$. The (piecewise constant) competing hazards for death or other cancer $\lambda_2 = \lambda_{2\_a} \cdot \beta_{2,country}$ and for hysterectomy $\lambda_3 = \lambda_{3,a,country}$ were computed without further covariates.

Absolute risk estimates over 5 years were calculated for each woman, based on her age and exposure profile at entry and on the estimates from the relative risk model and the age- and country-specific hazards model. Software written in SAS to compute absolute risk estimates for any age, projection length, and combination of risk factors is available for download **as online resource**.

Assessment of model quality
Five-fold cross-validation was applied to our modeling processes to obtain unbiased results on model quality on data portions that had been kept aside from model building steps. Discriminative quality was evaluated with an extension of the area under the ROC-curve statistic (AUROC), the concordance statistic 'Overall C' (18). The integrated discrimination improvement (IDI) index (19) was used to examine the improvement in discrimination between the epidemiologic risk model and a model based on age and country alone. Calibration of absolute

risk estimates was assessed as the ratio of expected to observed cases overall. The Hosmer-Lemeshow test (20) was used to evaluate model calibration over deciles of the risk score. All these analyses were stratified by country, and all were generated using SAS (9.3) software.

Results

The distribution of the risk factors in the study data and characteristics of incident cases of EC are presented in **Table 1**. The women's median age at recruitment was 50 years (95% were between age 27 and 68). The median follow-up of the study was 11.7 years (range: <1 - 16 years), and 855 incident cases of EC were diagnosed at median age at diagnosis of 61 years (2.5th-97.5th range: 46 – 76 years). Other incident cancers were diagnosed in 11,384 women, and 5,142 women died without developing cancer. At recruitment the women's BMI was on average 24 kg/m^2. Forty-five per cent (45%) of the women were premenopausal, 39% were postmenopausal at baseline. For postmenopausal women the median age at menopause was 50 years. The majority (59%) of women had used oral contraceptives (OC), and about 14% were former or current HRT users, 2% reported that they had diabetes, 17% hypertension and 13% hyperlipidemia. Eighteen percent were current smokers, 21% were former smokers and 60% had never smoked.

After backwards elimination, the following variables were selected into the prediction model: BMI (continuous), age at menopause and duration of HRT-use, each increasing the risk with increasing values; age at menarche, duration of OC use (years) by BMI category of lean (BMI < 25), overweight (BMI 25-30) or obese women (BMI>30), parity, age at first full term pregnancy, each decreasing the risk with increasing values, menopausal status and smoking status (smoking reduced risk among postmenopausal women). Hazard ratios (HR) are presented in **Table 2**. Overall, there was a five-fold difference in relative risk between the women in highest and lowest quintiles of the age-adjusted relative risk score (95%-CI: 3.86-6.56).

In terms of absolute risk estimates, **Figure 1** shows the age-specific percentiles of estimated five-year risk of EC. The 80th percentile of the women's five-year probability of developing EC varied from 0.23% at age 50-55 to 0.46% at age 65-70, and for women within the lowest 20% of risk the corresponding age-specific five-year probabilities were below 0.07% and 0.11%, respectively. For women with the worst risk score (i.e. highest BMI category, early age at first period, no OC-use, nulliparous, using HRT for 13 years and a non-smoker) the predicted five-year risk is 1.6% at age 50 and 4.2% at age 65. The full relative risk score based on the coefficients indicated in Table 2 is in the **Appendix**.

A risk model based on age as the only predictor showed a C-statistic of 71% (95%CI: 64-77%). The full epidemiologic risk model showed a C-statistic of 77% (95%CI: 68-85%), and provided a significantly better overall discrimination than the model based on age only, as indicated by an integrated discrimination improvement (IDI) index of 0.18% (95%CI:0.04-0.3%).

The absolute risk model predicted a total of 313 EC cases to occur in our cohort within five years after recruitment whereas 317 were observed during that time (expected-to-observed [E/O] ratio of 0.99). The Hosmer-Lemeshow test showed no significant deviation from the null hypothesis of good model fit (p= 0.08). Plots of observed _vs_. predicted risk by age and by country showed generally good internal calibration (**Figure 2**). Moreover, in each country the observed and predicted risk estimates from our cohort data showed a fair degree of concordance with age-specific incidence rates from regional cancer registries (21) (**Figure 2**).

Discussion

We developed an epidemiologic risk model for the prediction of women's individual probabilities of developing EC within a Western European population. Overall, our model provides significant discrimination between groups of women at high and low risk to develop EC, and its overall discrimination capacity (C-statistic) of about 77% is better than what is currently reported for prediction models of breast cancer (53-66%) (22, 23). Although a large part of the overall discrimination capacity is based on a woman's age (C= 71%), the information on the selected risk factors of BMI, reproductive history and exogenous hormone use does provide substantial improvement to the prediction of EC risk.

Only one other epidemiologic model for the prediction of absolute risk of EC has recently been developed by Pfeiffer et al. (24), using data from three large prospective cohort studies in the US. The risk factors in that model largely overlap with ours, although some differences also exist. Because our cohort includes a considerable proportion of younger women we could derive risk estimates for women from the age of 40. Our selected model also included interaction terms of BMI with OC use, and of smoking with menopausal status. Comparing Pfeiffer's model with ours on the eligible part (in terms of age-range and concurring variable definition) of EPIC cohort data resulted in an overall C-statistic of 66.8% (95%-CI: 63.5-70.2%) and 69.9% (95%-CI: 61.7-78.0%) respectively, indicating similar discrimination capacity for both models. However, with a ratio of expected to observed cases of 2.4, Pfeiffer's model was poorly calibrated to our cohort (on this data subset our model showed an E/O ratio of 0.98). Since that model had been developed and validated on US-cohort data, we would expect our model to perform equally

poorly on the US populations studied by Pfeiffer et al. The issue of model performance in the light of population variation is addressed in greater detail in **supplementary materials available as online resource**.

An important use of cancer risk prediction models in preventive oncology is the identification of high-risk individuals who may benefit from targeted interventions such as chemopreventive treatments or screening programs. By focusing these types of intervention on high-risk individuals it is often possible to improve the balance between benefits (prevention of cancer; improved chances of cure after early detection) and harms (e.g. treatment side-effects; false-positive screening diagnoses) (11). With regard to EC, one conceivable type of chemopreventive treatment is the use of progestins (7, 8, 10) in the form of combined OC (25), or progestin-based alternatives such as a depot of medroxyprogesterone acetate (MPA) (26) or a levonorgestrel-releasing intra-uterine system (LNG-IUS) (27).

Regarding EC screening, existing modalities include measurement of endometrial thickness by trans-vaginal ultrasound sonography (TVS) and concurrent endometrial biopsy (28). Recent observational data from a large-scale cohort in the United Kingdom provided some indication of the utility of targeted screening high-risk groups (12). Limiting TVS screening to a higher risk group selected on the basis of epidemiologic risk factors improved sensitivity and specificity to 84% and 90% respectively, and calculations showed that 40% of EC cases could have been diagnosed by screening only 25% of the population.

Whereas the prospective cohort design and the large size of the EPIC cohort and EC case numbers are major strengths, our study also has potential limitations. First of all, our study lacked information on several major risk factors, both at baseline and during prospective follow-up. This includes information on polycystic ovary syndrome, which among premenopausal women is associated with an approximate three to five-fold increase in EC risk before menopause (6, 7) and rates of hysterectomy during the follow-up period. Although our absolute risk models were adjusted for estimates of age- and country-specific rates of hysterectomy, the lack of incident hysterectomy data at the individual level has likely led to some reduction in the discriminating capacity of our model. A second possible limitation of our study is that we did not have complete information on the types of HRT used. Furthermore, we had no complete information on histological classification of endometrial tumors. Our prediction model likely estimates risk predominantly for the development of type-I (endometrioid) tumors, which represent about 85% of all EC occurrences.

External validation of our model on independent data remains open. However, both model building and estimation were separated from validation via cross-validation. Our model showed

good internal calibration of absolute risk estimates as compared to observed incidence rates within the EPIC cohorts. Comparisons of cohort-specific estimates with incidence rates from local cancer registries showed modest discrepancies in some of the EPIC countries (**Figure 2**). In a detailed examination of model performance in EPIC subcohorts we have estimated the variation of discrimination and calibration in different European populations (**online resource**). We find a strong heterogeneity between populations originating from different countries in crude incidence rates and in the country-effects estimated in our parametric model, which are adjusted for the EC-risk factors. Thus, a calibrating constant appears to be necessary to achieve good model validity for each individual country. External validation in a single external cohort may not be sufficient to assess general validity of such an EC-risk model, as can also be seen from the results of Pfeiffer's model applied to our cohorts. Hence we propose that in general the question of model calibration may be best assessed by estimating heterogeneity of model performance in diverse cohort populations.

EPIC cohort participants may not perfectly represent the population at large. This does not put the validity of our risk model into question, but it does restrict interpretation of our data with regard to the exact percentile distribution of women in the general population that may reach given absolute risk thresholds.

This epidemiologic risk model could be extended using information on possible (pre-) clinical symptoms, in particular postmenopausal vaginal bleeding symptoms, previous breast cancer and tamoxifen use (29), factors that are all likely to further discriminate women by the likelihood of developing, or already having EC. Another possible extension is the additional use of serum-based biomarkers of EC risk, such as endogenous hormones (30, 31) or biomarkers for early tumor detection (32).

As our model seems well calibrated with good discriminative capacity, we propose it could be used as first step in a multilevel prevention procedure and thus builds an important resource for further research on the prevention of EC in the general population in Western Europe. The observed population variation in EC-risk and EC-risk model performance suggests, that a large-scale world-wide cohort-consortium approach would be needed to assess and adjust for the population variation, to achieve an EC-risk model with general validity.

Appendix

The relative risk score as part of our model can be derived from the parameter estimates of the relative risk estimates (log(RR)) presented in Table 2 with the following formula:

$RR = exp$ [0.030 x **BMI** – 0.023 x (**age at 1st period** – 13) – 0.019 (if **lean OC-user**) – 0.013 (if **overweight OC-user**) – 0.036 (if **obese OC-user**) - 0.023 x **duration of OC-use** (in years) – 0.051 (if **single parous**) – 0.10 (if **2 full-term pregnancies**) – 0.22 (if **3 or more full-term pregnancies**) – 0.017 x (**age at 1st full-term pregnancy** – 24) – 0.088 (if **perimeno-pausal**) – 0.20 (if **postmenopausal**) + 0.029 x (**age at menopause** – 50) + 0.031 x **duration of HRT-use** (in years) – 0.11 (if **premenopausal former smoker**) + 0.040 (if **premenopausal current smoker**) – 0.12 (if **postmenopausal former smoker**) – 0.21 (if **postmenopausal current smoker**) – 0.14 (if **perimenopausal former smoker**)].

Funding

Table 1: Baseline characteristics of cohort subjects included in the present analyses.

		Total		EC-Cases	
Women included in analyses:	*N*	**201811**		**855**	
Age at recruitment	Median (2.5 -97.5%)	50	(27; 68)	55	(41; 69)
Duration of follow-up (years)	Median (min - max)	11.68	(0.01; 16.29)	6.46	(0.05; 14.03)
Non-case	*N* (%)	184,430	(91%)		
Other incident cancer	*N* (%)	11,384	(6%)		
Death	*N* (%)	5142	(3%)		
Age at censoring or diagnosis	Median (2.5 -97.5%)	60.92	(37.50; 79.26)	61.37	(46.38; 76.17)
BMI	Median (min - max)	24.13	(12.97; 74.46)	26.39	(15.18; 58.74)
Age at 1st period	Median (min - max)	13	(8; 20)	13	(9; 18)
Ever full-term pregnancy	*N* (%)	165,859	(82%)	714	(84%)
Age at 1st full-term pregnancy	Median (min - max)	24	(12; 51)	24	(15; 40)
number of children	Median (2.5 -97.5%)	2	(1; 5)	2	(1; 4)
Premenopausal	*N* (%)	90,476	(45%)	207	(24%)
Perimenopausal	*N* (%)	32,070	(16%)	146	(17%)
Postmenopausal	*N* (%)	79,265	(39%)	502	(59%)
Age at menopause	Median (2.5 -97.5%)	50	(40; 57)	51	(40; 59)
Ever used OC	*N* (%)	119,662	(59%)	334	(39%)
Among users: duration of use (years)	Median (min - max)	5	(0; 25)	2	(1; 25)
Ever use of HRT	*N* (%)	29,232	(14%)	213	(25%)
Among users: duration of use (years)	Median (2.5 -97.5%)	1.50	(0.08; 13)	2.00	(0.08; 20)
Prevalent diabetes	*N* (%)	4706	(2%)	37	(4%)
Unknown	*N* (%)	3710	(2%)	14	(2%)
Prevalent hypertension	*N* (%)	33,372	(17%)	217	(25%)
Unknown	*N* (%)	20,452	(10%)	133	(16%)
Prevalent hyperlipidemia	*N* (%)	25,833	(13%)	139	(16%)
Unknown	*N* (%)	19,896	(10%)	126	(15%)
Smoking status Never	*N* (%)	120,419	(60%)	578	(68%)
Former	*N* (%)	42,352	(21%)	153	(18%)
Smoker	*N* (%)	35,920	(18%)	113	(13%)
Unknown	*N* (%)	3120	(2%)	11	(1%)
Highest educational level					
None	*N* (%)	12,662	(6%)	62	(7%)
Primary school completed	*N* (%)	47,851	(24%)	262	(31%)
Technical/professional school	*N* (%)	37,184	(18%)	156	(18%)
Secondary school	*N* (%)	46,037	(23%)	191	(22%)
Longer education (incl. University deg.)	*N* (%)	50,705	(25%)	149	(17%)
Not specified	*N* (%)	7372	(4%)	35	(4%)
Alcohol consumption (g/d)					
0	*N* (%)	33,809	(17%)	167	(20%)
>0-6	*N* (%)	83,459	(41%)	351	(41%)
>6-12	*N* (%)	35,000	(17%)	147	(17%)
>12-24	*N* (%)	29,847	(15%)	115	(13%)
>24-60	*N* (%)	18,161	(9%)	71	(8%)
>60-96	*N* (%)	1387	(1%)	3	(0%)
>96	*N* (%)	148	(0%)	1	(0%)
Physical activity index					
Inactive	*N* (%)	49,865	(25%)	224	(26%)
Moderately inactive	*N* (%)	71,918	(36%)	311	(36%)
Moderately active	*N* (%)	48,844	(24%)	183	(21%)
Active	*N* (%)	31,184	(15%)	137	(16%)

Table 2: Multivariate relative risk estimates (adjusted for age and center) for epidemiological factors in the risk prediction model (with 95%-confidence interval limits).

	Relative Risk	95% confidence interval
BMI per unit (kg/m²) increase	1.07	(1.06; 1.09)
Age at 1st period (per year, centered at med. 13)	0.95	(0.91; 0.99)
OC-use (ever) in lean women (BMI<25)	0.96	(0.79; 1.16)
OC-use (ever) in overweight women (25<BMI<30)	0.74	(0.48; 1.14)
OC-use (ever) in obese women (BMI>30)	0.92	(0.57; 1.48)
Duration of OC-use (per year)	0.95	(0.93; 0.96)
Parity: nulliparous	1.00	
one child	0.89	(0.72; 1.10)
two children	0.80	(0.67; 0.95)
three or more children	0.60	(0.50; 0.73)
Age at 1st full term pregn. (per year, centered at median 24)	0.96	(0.95; 0.98)
Premenopausal	1.00	
Perimenopausal	0.82	(0.63; 1.06)
Postmenopausal	0.63	(0.48; 0.84)
Age at menopause (per year, centered at median 50)	1.07	(1.05; 1.09)
Duration of HRT use (per year)	1.07	(1.06; 1.10)
Smoking status: never (ref)	1.00	
Smoking in premenopausal women:		
former	0.78	(0.56; 1.09)
Smoker	1.10	(0.80; 1.51)
unknown	1.45	(0.61; 3.48)
Smoking in postmenopausal women		
Former	0.77	(0.38; 1.56)
Smoker	0.62	(0.30; 1.30)
Unknown	0.44	(0.04; 4.69)
Smoking in perimenopausal women		
Former	0.73	(0.32; 1.68)
Smoker	1.00	(0.44; 2.27)
Unknown	1.33	(0.17; 10.23)

Figure 1: Age-group-specific 5th, 20th, 50th, 80th and 95th percentiles of estimated 5-year risk of developing endometrial cancer in the EPIC-cohort.
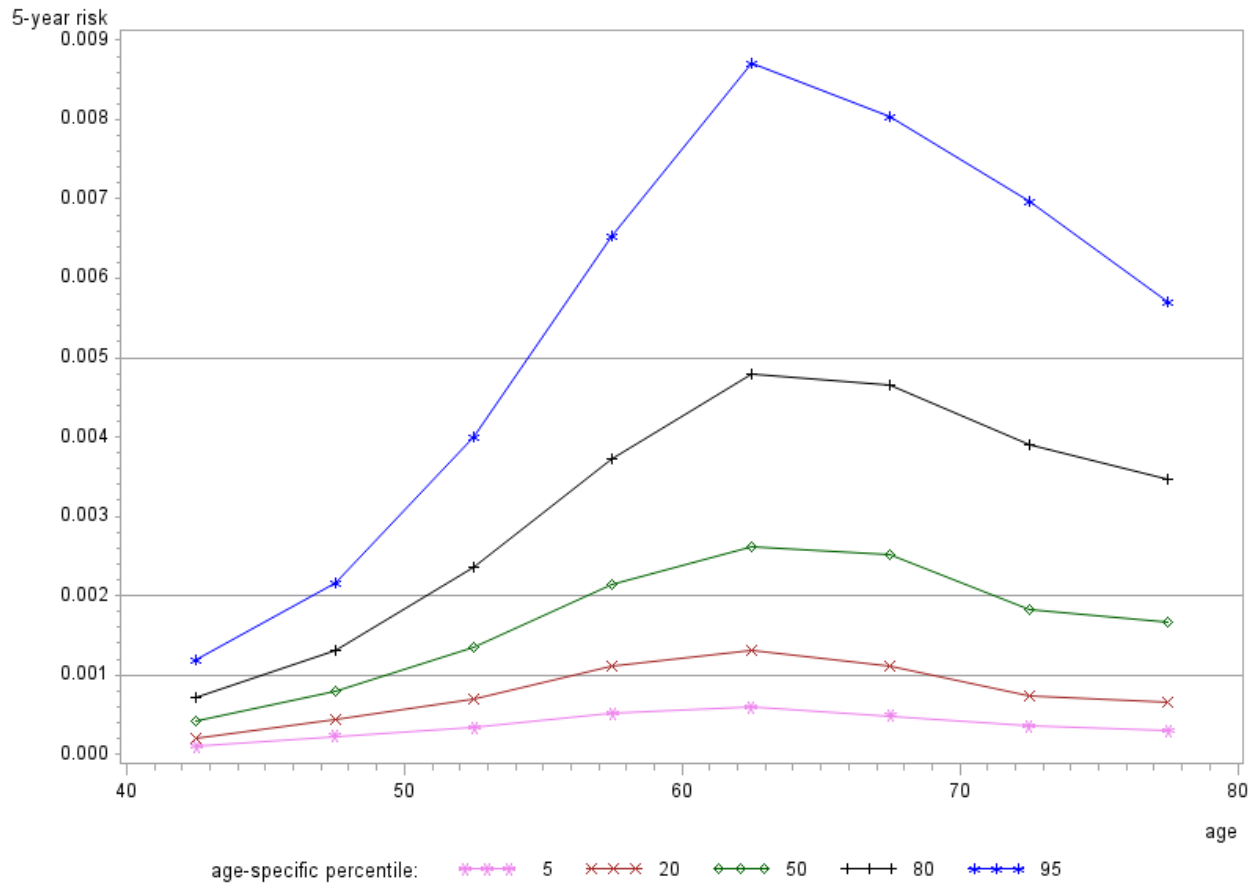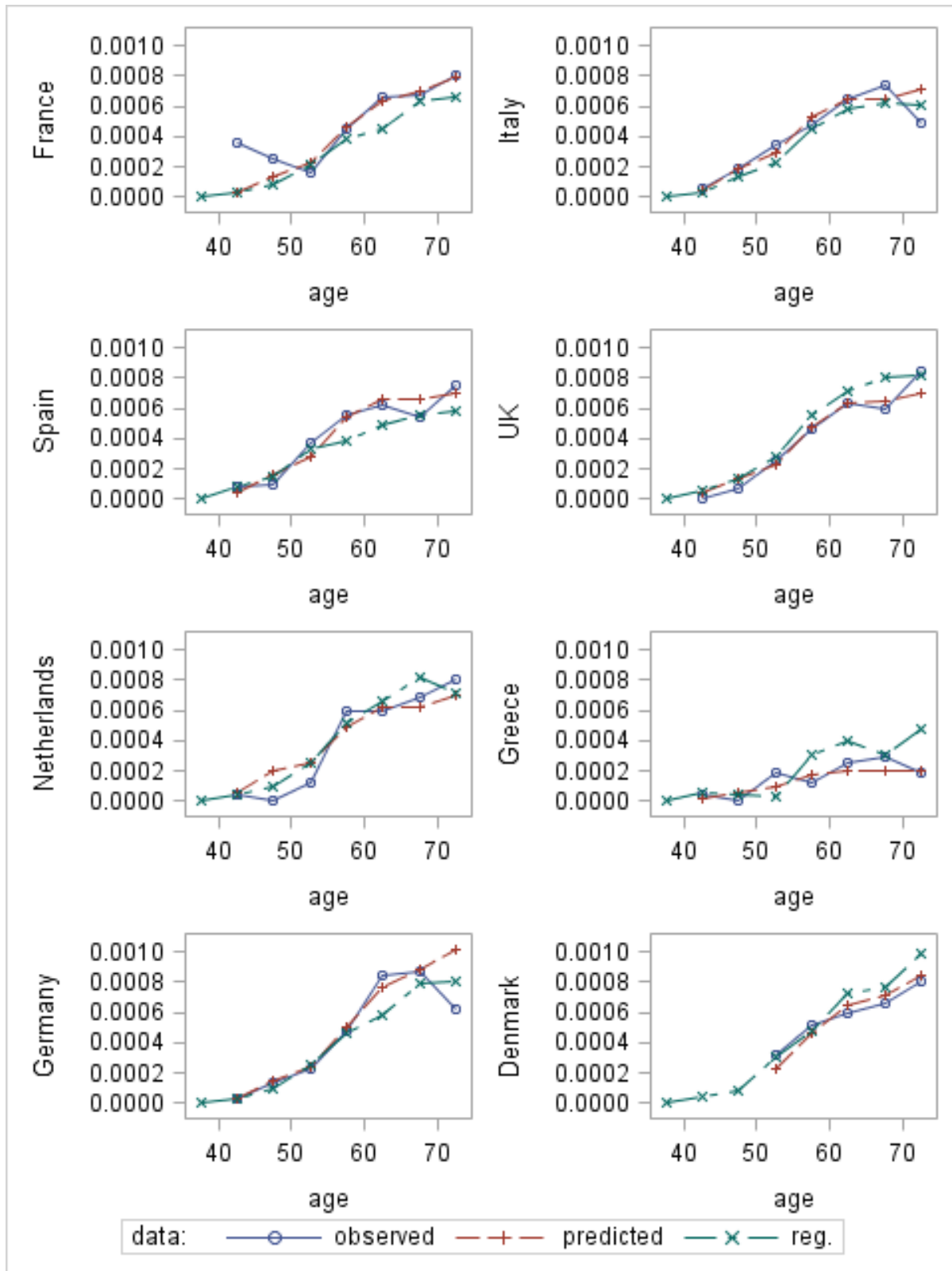
Figure 2: Country and age-specific annual EC risk as observed and predicted in our cohort, compared with rates from registries.

1. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, et al. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. Eur.J.Cancer. 2013;49(6):1374-403. doi:S0959-8049(13)00007-5 [pii];10.1016/j.ejca.2012.12.027 [doi]

2. Renehan AG, Tyson M, Egger M, Heller RF, Zwahlen M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. Lancet. 2008;371(9612):569-78. doi:S0140-6736(08)60269-X [pii];10.1016/S0140-6736(08)60269-X [doi]

3. Beral V, Bull D, Reeves G. Endometrial cancer and hormone-replacement therapy in the Million Women Study. Lancet. 2005;365(9470):1543-51. doi:S0140-6736(05)66455-0 [pii];10.1016/S0140-6736(05)66455-0 [doi]

4. Allen NE, Tsilidis KK, Key TJ, et al. Menopausal hormone therapy and risk of endometrial carcinoma among postmenopausal women in the European Prospective Investigation Into Cancer and Nutrition. Am.J.Epidemiol. 2010;172(12):1394-403. doi:kwq300 [pii];10.1093/aje/kwq300 [doi]

5. Dossus L, Allen N, Kaaks R, et al. Reproductive risk factors and endometrial cancer: the European Prospective Investigation into Cancer and Nutrition. Int.J.Cancer. 2010;127(2):442-51. doi:10.1002/ijc.25050 [doi]

6. Haoula Z, Salman M, Atiomo W. Evaluating the association between endometrial cancer and polycystic ovary syndrome. Hum.Reprod. 2012;27(5):1327-31. doi:des042 [pii];10.1093/humrep/des042 [doi]

7. Kaaks R, Lukanova A, Kurzer MS. Obesity, endogenous hormones, and endometrial cancer risk: a synthetic review. Cancer Epidemiol.Biomarkers Prev. 2002;11(12):1531-43.

8. Taylor HS, Manson JE. Update in hormone therapy use in menopause. J.Clin.Endocrinol.Metab. 2011;96(2):255-64. doi:96/2/255 [pii];10.1210/jc.2010-0536 [doi]

9. Lacey JV, Jr., Chia VM, Rush BB, et al. Incidence rates of endometrial hyperplasia, endometrial cancer and hysterectomy from 1980 to 2003 within a large prepaid health plan. Int.J.Cancer. 2012;131(8):1921-9. doi:10.1002/ijc.27457 [doi]

10. Yang S, Thiel KW, Leslie KK. Progesterone: the ultimate endometrial tumor suppressor. Trends Endocrinol.Metab. 2011;22(4):145-52. doi:S1043-2760(11)00014-2 [pii];10.1016/j.tem.2011.01.005 [doi]

11. Gail MH. Personalized estimates of breast cancer risk in clinical practice and public health. Stat.Med. 2011;30(10):1090-104. doi:10.1002/sim.4187 [doi]

12. Jacobs I, Gentry-Maharaj A, Burnell M, et al. Sensitivity of transvaginal ultrasound screening for endometrial cancer in postmenopausal women: a case-control study within the UKCTOCS cohort. Lancet Oncol. 2011;12(1):38-48. doi:S1470-2045(10)70268-0 [pii];10.1016/S1470-2045(10)70268-0 [doi]

13. Friedenreich C, Cust A, Lahmann PH, et al. Anthropometric factors and risk of endometrial cancer: the European prospective investigation into cancer and nutrition. Cancer Causes Control. 2007;18(4):399-413. doi:10.1007/s10552-006-0113-8 [doi]

14. Friedenreich C, Cust A, Lahmann PH, et al. Physical activity and risk of endometrial cancer: the European prospective investigation into cancer and nutrition. Int.J.Cancer. 2007;121(2):347-55. doi:10.1002/ijc.22676 [doi]

15. Al-Zoughool M, Dossus L, Kaaks R, et al. Risk of endometrial cancer in relationship to cigarette smoking: results from the EPIC study. Int.J.Cancer. 2007;121(12):2741-7. doi:10.1002/ijc.22990 [doi]

16. Riboli E, Hunt KJ, Slimani N, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. Public Health Nutr. 2002;5(6B):1113-24.

17. Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. Stat.Med. 2007;26(11):2389-430. doi:10.1002/sim.2712 [doi]

18. Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. Stat.Med. 2004;23(13):2109-23. doi:10.1002/sim.1802 [doi]

19. Pencina M, D'Agostino R, D'Agostino R, Vasan R. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Statistics in medicine. 2008;27(2):157-72.

20. David W.Hosmer J, PhD, Stanley Lemeshow P. Applied Logistic Regression. 2nd ed. New York: John Wiley & Sons; 2000.

21. Ferlay J SH, Bray F,Forman D,Mathers C and Parkin DM. Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10 [Internet]. http://globocan.iarc.fr/ 2008.

22. Anothaisintawee T, Teerawattananon Y, Wiratkapun C, Kasamesup V, Thakkinstian A. Risk prediction models of breast cancer: a systematic review of model performances. Breast Cancer Res.Treat. 2012;133(1):1-10. doi:10.1007/s10549-011-1853-z [doi]

23. Husing 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):601-8. doi:jmedgenet-2011-100716 [pii];10.1136/jmedgenet-2011-100716 [doi]

24. 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. doi:10.1371/journal.pmed.1001492 [doi];PMEDICINE-D-13-00139 [pii]

25. Cibula D, Gompel A, Mueck AO, et al. Hormonal contraception and risk of cancer. Hum.Reprod.Update. 2010;16(6):631-50. doi:dmq022 [pii];10.1093/humupd/dmq022 [doi]

26. Lu KH, Loose DS, Yates MS, et al. Prospective multicenter randomized intermediate biomarker study of oral contraceptive versus depo-provera for prevention of endometrial cancer in women with Lynch syndrome. Cancer Prev.Res.(Phila). 2013;6(8):774-81. doi:1940-6207.CAPR-13-0020 [pii];10.1158/1940-6207.CAPR-13-0020 [doi]

27. Chin J, Konje JC, Hickey M. Levonorgestrel intrauterine system for endometrial protection in women with breast cancer on adjuvant tamoxifen. Cochrane.Database.Syst.Rev. 2009(4):CD007245. doi:10.1002/14651858.CD007245.pub2 [doi]

28. Breijer MC, Peeters JA, Opmeer BC, et al. Capacity of endometrial thickness measurement to diagnose endometrial carcinoma in asymptomatic postmenopausal women: a systematic review and meta-analysis. Ultrasound Obstet.Gynecol. 2012;40(6):621-9. doi:10.1002/uog.12306 [doi]

29. Burbos N, Musonda P, Giarenis I, et al. Predicting the risk of endometrial cancer in postmenopausal women presenting with vaginal bleeding: the Norwich DEFAB risk assessment tool. Br.J.Cancer. 2010;102(8):1201-6. doi:6605620 [pii];10.1038/sj.bjc.6605620 [doi]

30. Allen NE, Key TJ, Dossus L, et al. Endogenous sex hormones and endometrial cancer risk in women in the European Prospective Investigation into Cancer and Nutrition (EPIC). Endocr.Relat Cancer. 2008;15(2):485-97. doi:15/2/485 [pii];10.1677/ERC-07-0064 [doi]

31. Dossus L, Lukanova A, Rinaldi S, et al. Hormonal, metabolic, and inflammatory profiles and endometrial cancer risk within the EPIC cohort--a factor analysis. Am.J.Epidemiol. 2013;177(8):787-99.

32. Angioli R, Capriglione S, Aloisi A, et al. REM (risk of endometrial malignancy): a proposal for a new scoring system to evaluate risk of endometrial malignancy. Clin.Cancer Res. 2013;19(20):5733-9.

Supplementary material to manuscript on

An epidemiological model for prediction of endometrial cancer risk in Europe

Corresponding author:
Anika Hüsing,
German Cancer Research Center (DKFZ), Division of Cancer Epidemiology (C020)
Im Neuenheimer Feld 581, D-69120 Heidelberg
Phone: +49-(0)6221-42-2214,
Fax: +49-(0)6221-42-2203
E-Mail: a.huesing@dkfz.de

Generalizability of model performance

1. Comparison between countries:

International registry data on endometrial cancer show country-wise differences in (age-specific) incidence rates as presented in Figure S1. The overall trend is an increase of risk from the age of 48 to its 5-6-fold level between age 60 and 65 and then levelling or slight decline.
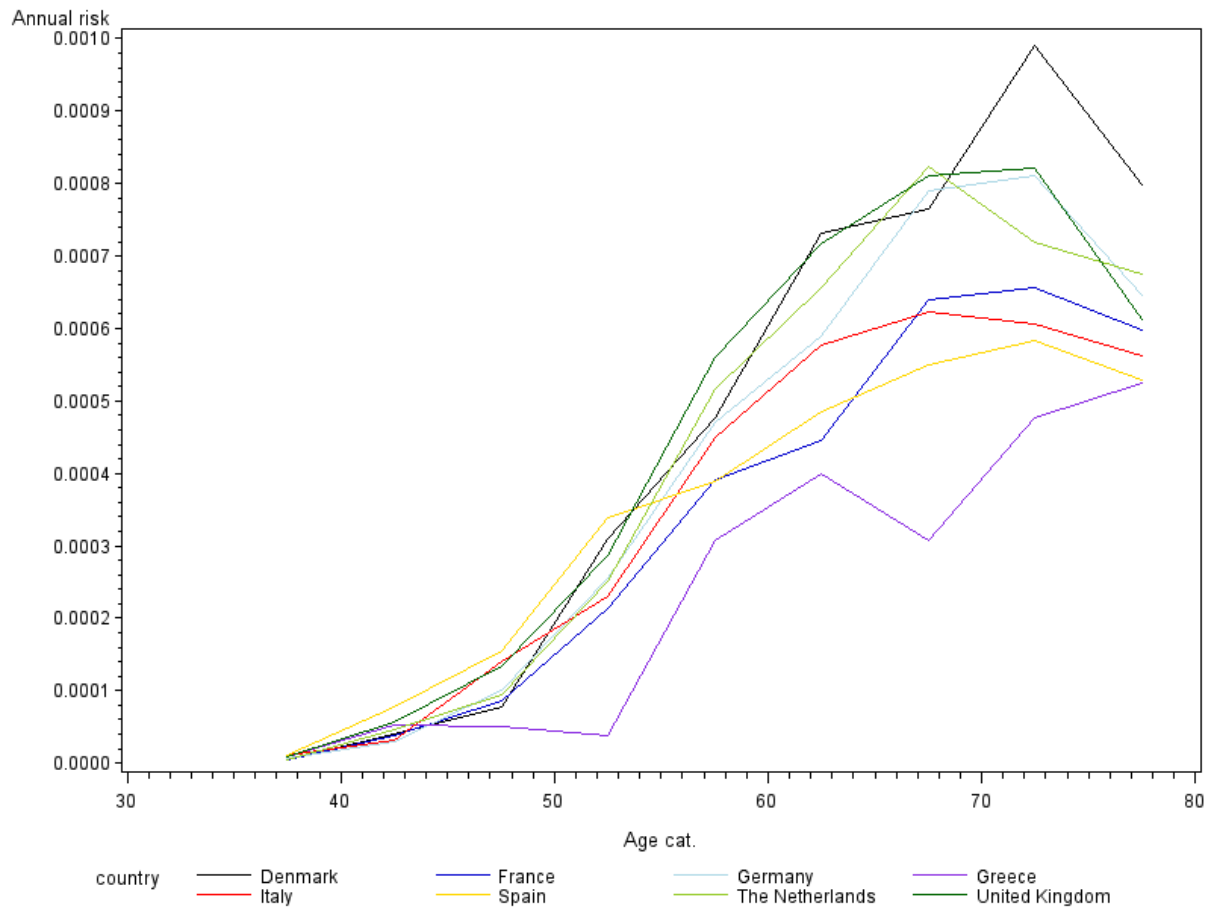
Figure S1: annual incidence rates of endometrial cancer in Western Europe in 5-year age-groups as provided by registries.

For our model we use age as the underlying time-scale. For derivation of the relative risk score, center is used as stratification factor, to account for differences due to study-setup and investigator. While it makes sense to rule out inter- and intra-country differences due to different study situations in estimating general relative risks, a center-specific differentiation would not make sense for estimation and application of absolute risks in a clinical setting. The final model includes simple categorical effects (in a proportional parametric survival model) for country to allow for variation between populations in terms of average effects. These estimated country-effects are per construction adjusted for the common effect of the covariates selected into the final model and for age – thus, the country-effects describe population specific heterogeneity in risks in a uniform manner that is not accounted for by the selected, risk-related covariates. These country-related model effects may reflect uncaptured country-specific variation in life-style factors such as diet, physical activity, family life or health care provision.

We estimate the absolute risk as cumulative incidence according to formula 1:

$$\pi(a, b, x) = F(a, b, x) = \int_a^b \lambda_1(x) \cdot \exp(s \cdot (-\lambda_1(x) - \lambda_2 - \lambda_3)) \, ds,$$

$$F(a, b, x) = \int_a^b \lambda_1(x) \cdot \exp(s \cdot (-\lambda_1(x) - \lambda_2 - \lambda_3)) \, ds,$$

where $\lambda_1(x) = \lambda_{1\_a} \cdot \beta_{1,country} \cdot rr(x)$ represents the cause specific risk of EC, and is calculated as product of (piecewise constant) baseline hazard λ_{1_a} in age-interval a, a constant factor $\beta_{1,country}$ to account for differences in incidence rates between countries, and the relative risk score derived from the Cox-model as $rr(x) = \exp(\beta_{rr}'x)$. The (piecewise constant) competing hazards for death or other cancer enter this calculation as $\lambda_2 = \lambda_{2\_a} \cdot \beta_{2,country}$ and for hysterectomy $\lambda_3 = \lambda_{3,a,country}$.

The EC-specific component resolves to

$$\pi(a, b, x) = 1 - \exp\left(-\exp\left(\log(b - a) - \beta_{rr} \cdot x - \lambda_{1\_a} - \beta_{1,country}\right)\right)$$

With the hazard being $h(t) = \exp\left(-\beta_{rr} \cdot x - \lambda_{1\_a} - \beta_{1,country}\right)$ country-specific hazard ratios can be derived as

$$HR(t) = \frac{h(t|country1)}{h(t|country\_ref)} = \frac{\exp(-\beta_{rr} \cdot x - \lambda_{1\_a} - \beta_{1,country1})}{\exp(-\beta_{rr} \cdot x - \lambda_{1\_a} - \beta_{1,country\_ref})} = \exp(\beta_{1,country\_ref} - \beta_{1,country1}).$$

For this analysis we ignore additional complexities due to competing hazards. For a closer look at country-specific rates, we derive estimates as presented in Table S1, comparing average incidence rates reported from registries with hazard ratios calculated from our model and from a model fit to our data with the same model assumptions, but without the relative risk score, i.e. without adjustment for covariates. These estimates should be more comparable to registry rates, which are also not adjusted for covariates. Differences between these estimates and registry rates may reflect country-specific deviations of our sample from the general population (i.e. incomplete representativeness). The order of presentation follows the ranking by age-standardized incidence rates as reported by EUCAN (1), using Denmark as reference category.

Table S1: Registry incidence rates and hazard ratios from our risk model and from a model without the common covariate effect.

	Ranking of registry incidence rates (age stdzd.)			Model without covariate risks, age-standardized (p<=0.0001)			Model including covariate risks and age (p<=0.0001)		
		IR*	IR relative to DK	HRlow	**HRatio**	HRup	HRlow	**HRatio**	HRup
1	UK	19.7	1.026	0.741	**0.886**	1.059	0.787	**0.941**	1.125
2	Denmark	19.2	1.000		**1.000**			**1.000**	
3	Italy	19.2	1.000	0.902	**1.070**	1.270	0.798	**0.948**	1.126
4	Netherlands	17.7	0.922	0.737	**0.897**	1.092	1.032	**1.259**	1.537
5	Germany	16.9	0.880	0.801	**0.989**	1.222	0.808	**0.998**	1.233
6	Spain	16.3	0.849	0.897	**1.079**	1.298	0.629	**0.757**	0.912
7	France	15.2	0.792	0.442	**0.534**	0.644	0.475	**0.573**	0.692
8	Greece	10	0.521	0.192	**0.270**	0.381	0.134	**0.189**	0.266

* Age Standardised Incidence Rate (European) per 100,000 as of 2012,

The number of cases per country is limited in our sample, which hampers our estimation power.

However, we do see a variation between countries, in line with the variation in national incidence rates. Greece has extremely low registry rates and appears as an outlier in our sample data. The latter effect may at least in part be ascribed to an ascertainment bias caused by the local complex active follow-up scheme.

2. Discrimination across countries:

In independent, and in our case neutral selection processes both the Pfeiffer group and we selected the same or at least very similar risk factors into the prediction model, and the risk factors were assigned similar weights.

The model provided by Pfeiffer is designed only for women of at least 50 years of age and up to a prediction age of 90.

We provide here also results from a model estimated according to our distributional assumptions and including our relative risk score, but without country parameters. Estimates from this model should be more in line with those from Pfeiffer's model, which also includes no country-component.

For our model (in cross-validated data) we find discrimination to differ by country. Because also the sample age-range affects discrimination, and age-ranges differ between countries, we use only women who were between 50 and 70 years old at recruitment for a standardized comparison of discriminative capacity between countries.

Table S2: Discrimination in terms of overall C (with 95%-confidence limits) of 5-year risk prediction in 50-70 year old women in our sample provided by our model and by Pfeiffer's, and by a model with our relative risk score, without a country-effect:

		full model			our model without country effect			Pfeiffer's model		
	Cases in 5 years	C	Low 95%ci C	Up 95%ci C	C	Low 95%ci C	Up 95%ci C	C	Low 95%ci C	Up 95%ci C
All across	253	**0.683**	0.650	0.717	**0.653**	0.619	0.688	**0.670**	0.636	0.705
France	45	**0.666**	0.590	0.742	**0.614**	0.532	0.696	**0.647**	0.563	0.731
Italy	39	**0.631**	0.541	0.721	**0.627**	0.542	0.711	**0.639**	0.560	0.717
Spain	24	**0.694**	0.579	0.809	**0.691**	0.579	0.804	**0.626**	0.513	0.739
UK	31	**0.736**	0.648	0.824	**0.688**	0.590	0.786	**0.728**	0.634	0.821
The Netherlands	26	**0.722**	0.626	0.817	**0.674**	0.566	0.782	**0.710**	0.601	0.819
Greece	12	**0.692**	0.561	0.822	**0.628**	0.487	0.770	**0.587**	0.427	0.748
Germany	26	**0.762**	0.665	0.860	**0.735**	0.630	0.840	**0.741**	0.627	0.855
Denmark	50	**0.658**	0.582	0.734	**0.632**	0.554	0.711	**0.680**	0.598	0.761

Discrimination is overall similar for Pfeiffer's model and ours. Both models reveal a varying discriminative capacity between countries, even within this constrained age-range. Within countries the models' discriminative capacity appears strongly correlated. .

Comparison of our two model results demonstrates that the country-component improves the discriminative capacity not only between countries, but due to its proportionality on risk also within countries. With an increase in C (Delta-C) of 3% this would commonly be considered as relevant.

3. Calibration across countries:

Pfeiffer's model was externally calibrated but performed poorly in our sample. However, since that model was well calibrated to the (white) US population and performed reasonably well in validation in this population, we would assume that our model, which is calibrated towards specific Western European countries, would perform equally poorly for the populations studied

by Pfeiffer et al. for their model development. Our own E/O ratios (as calculated from cross-validation results) also show a certain variation, with best fit for Denmark, the Netherlands, Germany and marked underestimation for the Italy/Greece sample, which was joined for the cross-validation analysis, because of the small size of the greek sample. This underestimation may be another result of the biased case-ascertainment in Greece, as mentioned above.

Table S3: Ratios of expected to observed numbers of cases in countries and overall, together with 95%-confidence intervals calculated from our final model, our model without country-effects and Pfeiffer's model.

	Full model			Model w/out country-effect			Pfeiffer's model		
	E/O	low 95% ci	up 95% ci	E/O	low 95% ci	up 95% ci	E/O	low 95%ci	up 95%ci
Denmark	**1.01**	0.76	1.33	**1.38**	1.04	1.82	**2.14**	1.62	2.82
France	**1.06**	0.79	1.42	**1.52**	1.13	2.03	**2.29**	1.70	3.07
Italy&Greece	**0.79**	0.60	1.04	**1.76**	1.34	2.31	**2.97**	2.25	3.90
Spain	**0.90**	0.60	1.34	**1.71**	1.15	2.55	**3.17**	2.13	4.73
UK	**0.91**	0.64	1.30	**1.45**	1.02	2.07	**2.11**	1.49	3.00
The Netherlands	**1.01**	0.69	1.49	**1.28**	0.87	1.88	**1.98**	1.35	2.91
Germany	**0.99**	0.68	1.46	**1.22**	0.83	1.79	**2.11**	1.44	3.10
All	**0.95**	0.84	1.07	**1.49**	1.32	1.69	**2.41**	2.13	2.72

With our sample including data from eight different Western European countries, we can have a look at the variation of model performance for different populations.

Overall we find agreement on models' discriminative capacities, but a vulnerability concerning the models' calibration quality. This makes us wary about the recommendation of the application of risk models across populations in general terms.

We observe a North/South gradient in our country effects across Europe, with an overall agreement of expected and observed cases for the UK, Italy, Denmark, Germany, and The Netherlands, assigning Spain, France and especially Greece a special status.

Our final model provides reasonable discrimination and good calibration qualities for all countries in our sample.

4. In conclusion:

1. Incidence rates differ between countries both in our sample and according to registry rates.
2. These differences can in part be accounted for by differences in the distribution of epidemiological factors (life-style, diet, living-conditions, cultural differences etc.).
3. However, we still find a significant difference between countries remaining, when we estimate country specific absolute risk levels in the presence of our relative risk score.
4. Both calibration and discrimination show heterogeneity between countries.
5. Addition of a country effect to the model improves discrimination to a degree of difference in C (Delta-C) of 3%, which is usually considered as relevant and improves calibration (changing E/O from 1.47 overall to 0.97).

External validation provides no safeguard against miscalibration. Calibration seems to be more affected from variation between populations than discrimination. It appears that calibration may suffer severely when a model is applied to a new population. From our cross-population comparison we also deduce that calibration is an issue, which depends on differences between populations in factors which are not covered by model components. With a simple categorical effect for country (in our setting of proportional parametric survival models) we allow absolute risk levels to be on average different for populations from different countries to capture population heterogeneity not accounted for with our risk-factors. This heterogeneity may be related for example to diet, life-style, but also to infrastructure (child-care, mobility, diet), or health care systems, offering medication and diagnosis according to standards differing by country and also by time. Both calibration and discrimination are much improved with this country-factor.

Thus we conclude that for an EC-risk model general validity may be achieved only in a setting which provides enough power to consider such population variation. Considering the low incidence of this disease, a large-scale world-wide cohort consortium may be necessary to gather sufficient data from different populations. As seen by the performance of Pfeiffer's model, external validation on single cohorts provides no safeguard against miscalibration in principle.

1.   Steliarova-Foucher E OCM, Ferlay J,Masuyer E,Forman D,Comber H,Bray F. European Cancer Observatory: Cancer Incidence, Mortality, Prevalence and Survival in Europe. **International Agency for Research on Cancer**. 2012. http://eco.iarc.fr. Accessed 5/1/2014 2014.