Critical Reviews in Oncology / Hematology 126 (2018) 92–99

Contents lists available at ScienceDirect

Critical Reviews in Oncology / Hematology

journal homepage: www.elsevier.com/locate/critrevonc

Prediction models for endometrial cancer for the general population or symptomatic women: A systematic review

Maaike Alblas[a,*], Kimberley B. Velt[a], Nora Pashayan[b], Martin Widschwendter[c], Ewout W. Steyerberg[a,d], Yvonne Vergouwe[a]

[a] Department of Public Health, Erasmus MC, Rotterdam, The Netherlands
[b] Department of Applied Health Research, University College London, London, United Kingdom
[c] Department of Women's Cancer, University College London, London, United Kingdom
[d] Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands

ARTICLE INFO

Keywords:
Risk model
Diagnostic model
Endometrial cancer
Systematic review
Methods
Reporting

ABSTRACT

Objective: To provide an overview of prediction models for the risk of developing endometrial cancer in women of the general population or for the presence of endometrial cancer in symptomatic women.
Methods: We systematically searched the Embase and Pubmed database until September 2017 for relevant publications. We included studies describing the development, the external validation, or the updating of a multivariable model for predicting endometrial cancer in the general population or symptomatic women.
Results: Out of 2756 references screened, 14 studies were included. We found two prediction models for developing endometrial cancer in the general population (risk models) and one extension. Eight studies described the development of models for symptomatic women (diagnostic models), one comparison of the performance of two diagnostic models and two external validation. Sample size varied from 60 (10 with cancer) to 201,811 (855 with cancer) women. The age of the women was included as a predictor in almost all models. The risk models included epidemiological variables related to the reproductive history of women, hormone use, BMI, and smoking history. The diagnostic models also included clinical predictors, such as endometrial thickness and recurrent bleeding. The concordance statistic (*c*), assessing the discriminative ability, varied from 0.68 to 0.77 in the risk models and from 0.73 to 0.957 in the diagnostic models. Methodological information was often limited, especially on the handling of missing data, and the selection of predictors. One risk model and four diagnostic models were externally validated.
Conclusions: Only a few models have been developed to predict endometrial cancer in asymptomatic or symptomatic women. The usefulness of most models is unclear considering methodological shortcomings and lack of external validation. Future research should focus on external validation and extension with new predictors or biomarkers, such as genetic and epigenetic markers.

1. Introduction

Endometrial cancer is the sixth most common type of cancer in women worldwide and its incidence has been increasing since 1990 (Ferlay et al., 2013). This increase might be related to improvements in detection in the general population and in diagnostics in women with (postmenopausal) bleeding. Further, in many populations the body mass index (BMI) is rising and several studies have shown that adiposity is the strongest risk factor of endometrial cancer (Kyrgiou et al., 2017; Dixon, 2010; Collaboration NCDRF, 2016; Ng et al., 2014). Other risk factors that are associated with endometrial cancer are higher age, hypertension, diabetes, nulliparity, early menarche, late menopause, oestrogen uptake, and genomic alterations (MacMahon, 1974; Hecht and Mutter, 2006). Combining these risk factors in multivariable prediction models may help to identify women in the general population at high risk of developing endometrial cancer. Prediction models can also facilitate early diagnosis in symptomatic women.

Several risk and diagnostics models for endometrial cancer have been developed (Pfeiffer et al., 2013; Wong et al., 2016; Husing et al., 2016; Burbos et al., 2010; Giannella et al., 2014). The models can be used for risk prediction for prevention purposes. Particularly models with modifiable risk factors, such as BMI, hypertension, and oestrogen uptake may facilitate tailored preventive interventions on diet, lifestyle or drug use. This might reduce the incidence of endometrial cancer.

* Corresponding author at: Erasmus MC – University Medical Center Rotterdam, Department of Public Health Na-2401, P.O. 2040, 3000 CA Rotterdam, The Netherlands.
E-mail address: m.alblas@erasmusmc.nl (M. Alblas).

https://doi.org/10.1016/j.critrevonc.2018.03.023
Received 14 November 2017; Received in revised form 13 March 2018; Accepted 28 March 2018

Once endometrial cancer has developed, diagnostic models can be used for early diagnosis. Postmenopausal bleeding and increasing endometrial thickness are the most common symptoms of endometrial cancer and are often considered in these diagnostic models (Gull et al., 2003). The diagnostic models facilitate early diagnosis, which may result in efficient use of diagnostic resources and improved survival.

Since no overview of these models has been published so far, we aimed to systematically review multivariable models predicting the risk of endometrial cancer in the general population. We also systematically reviewed models for the presence of endometrial cancer in symptomatic women. We describe the model development, the included predictors, the predicted outcome, and any attempts to external validation to assess the quality of the models and determine if these models are ready for use in practice.

2. Methods

2.1. Search strategy

The search strategy that was used in this review was based on previous published searches (Damen et al., 2016; Ingui and Rogers, 2001) and other systematic reviews of prediction models (Smit et al., 2015; Meads et al., 2012; Mushkudiani et al., 2008). Specific terms for endometrial cancer were added to the search strategy. The index terms of papers that were considered relevant were manually searched to check if any search terms were missing from the search strategy. The final strategy (S1) was used in the PubMed and Embase databases in August 2017.

2.2. Inclusion criteria

We included all papers with the main aim of developing, validating or updating a model predicting the risk of endometrial cancer in the general population or presence in symptomatic women. Any multivariable (at least two predictors) prediction model was eligible for inclusion, including prediction scores or prediction tools. Only papers written in the English language were included. There was no restriction on publication date.

2.3. Screening process and data extraction

Two authors performed the screening process and data extraction. One author (MA) reviewed the titles and abstracts of all papers that were identified during the search, after which a random sample of 10% was checked by another author (KV). Both authors independently screened the full text of the remaining papers for eligibility. Disagreements were solved by discussion between the authors or consulting a senior author (YV).

The data extraction sheet was based on the CHARMS checklist. The data extraction sheet was pilot tested on two articles to ensure consistency between both authors. Subsequently, both authors performed the data extraction on all included papers. Specific attention was paid to four main topics (study design and methods, outcome and predictors, model development, model performance and model validation) of the CHARMS checklist, as these topics mainly influence the validity of the models.

Study design and methods: We identified the study design (e.g. case-control, cohort, case-cohort), source of data (e.g. hospital based or national registries) and size of the study population. In addition, the inclusion criteria for each study were assessed.

Outcome and predictors: We assessed the measurement and definition of both the outcome and predictors, and the handling of predictors (e.g. predictors were kept continuous or were dichotomized).

Model development: We assessed the following topics: handling of predictors, number of events per variable (EPV), number and handling of missing data (e.g. single imputation, multiple imputation), methods for selection of predictors in the multivariable model (e.g. univariate analyses or subject matter knowledge) and during multivariable modelling (backward or forward selection), modelling method (e.g. logistic regression, cox proportional hazards), shrinkage (e.g. penalized shrinkage or lasso) and model presentation (e.g. regression formula, score chart, nomogram or risk score).

Model performance and validation: Aspects concerning model performance and validation that were assessed were discrimination, calibration, internal validation (e.g. split-sample approach, cross-validation, bootstrapping) and external validation (e.g. geographical or temporal validation). Furthermore, the sensitivity, specificity, positive and negative likelihood ratios, and positive and negative predictive values of the diagnostic models were included in this topic, if reported.

3. Results

We identified 2756 papers during the initial search. These records were screened on title and abstract after which 23 records were included for full text screening. The low sensitivity of the search (less than 1% of the initial search result was included for full text screening) is in line with other searches, as a consequence of the lack of adequate search terms for prediction models. After full text screening, 9 papers were eligible for inclusion. In addition, 5 extra papers were identified by hand search, leading to the inclusion of 14 papers in this review (Fig. 1). Two papers developed prediction models for the general population (risk models), eight papers developed prediction models for symptomatic women (diagnostic models), one paper internally evaluated a model, two papers described the external validation of previous developed models and one paper described the extension of an existing prediction model.

3.1. Prediction models for endometrial cancer in the general population (risk models)

3.1.1. Study designs and population

The two studies that developed risk models used data from population based cohorts; one study used the European EPIC cohort (Husing et al., 2016) and one study used a cohort from the United States (Pfeiffer et al., 2013) (US) (Table 1). The data was collected using a

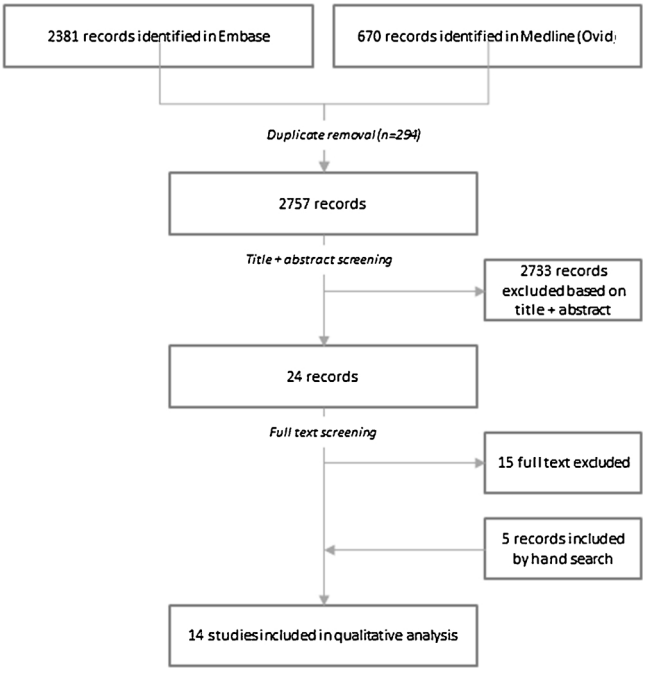

Fig. 1. Flow diagram of study selection process.

Table 1
Characteristics of 14 included studies.

First author	Year	Study design	Recruitment period	Total number of participants (number of cases)	Inclusion criteria	Aim of study
Risk models						
Fortner	2017	Nested case-cohort	1992-2000	716 (247)	Women without hysterectomy and prevalent cancer	Extension
Hüsing	2015	Prospective cohort	1992-2000	201,811 (855)	Women without hysterectomy and prevalent cancer	Development
Pfeiffer	2013	Prospective cohort	PLCO: 1993-2001 NIH-AARP: 1995-1996 NHS: 1990-2004	41,694 (462) 104,985 (1097) 37,241 (532)	Non-Hispanic white women aged 50 +	Development + validation
Diagnostic models						
Madkour	2017	Prospective cohort	Unknown	60 (10)	Women with postmenopausal bleeding	Development
Plotti	2017	Prospective cohort	2013-2016	298 (102)	Women aged 30 to 80 years with ultrasound endometrial abnormalities and scheduled hysteroscopy	Validation
Wong	2016	Retrospective cohort	2002-2013	4383 (168)	Women with vaginal bleeding	Development
Sladkevicius	2016	Prospective cohort	2009-2014	350 (80)	Postmenopausal women presenting with vaginal bleeding, ET ≥ 4.5 mm, and no fluid in the uterine cavity	Validation
Giannella	2014	Prospective observational	2008-2013	624 (72)	Postmenopausal women presenting with vaginal bleeding and ET ≥ 4 mm undergoing diagnostic hysteroscopy	Development
Angioli	2013	Prospective cohort	2010-2012	675 (88)	Women aged 40 to 65 years with ultrasound endometrial abnormalities and scheduled hysteroscopy	Development
Opolskiene	2011	Prospective cohort	2002-2009	261 (63)	Postmenopausal women presenting with vaginal bleeding, ET ≥ 4.5 mm and no fluid in the uterine cavity	Development
Burbos	2011	Prospective cohort	2006-2009	3347 (201)	Postmenopausal women presenting with vaginal bleeding	Development
Musonda	2011	Prospective cohort	2006-2009	3795 (221)	Postmenopausal women presenting with vaginal bleeding	Comparison
Burbos	2010	Prospective cohort	2006-2009	3047 (149)	Postmenopausal women presenting with vaginal bleeding	Development
Weber	1999	Case-control	1993-1995	194 (57)	Perimenopausal or postmenopausal women presenting with vaginal bleeding who had endometrial sampling	Development

Table 2
Overview of predictors included in risk models and diagnostic models for endometrial cancer.

	Age	Age at menopause	Age at menarche	BMI	(Duration of) HRT use	(Duration of) OC use	Parity	Age at FFTP	Menopausal status	Smoking	Biomarkers[2]	Recurrent bleeding
Risk models												
Fortner et al. (2017)	X	X	X	X	X	X	X	X	X	X	X	
Husing et al. (2016)	X	X	X	X	X	X	X	X	X	X		
Pfeiffer et al. (2013)	X	X		X	X	X	X		X	X		
Diagnostic models												
Madkour (2017)												
Wong et al. (2016)	XX	X		X			XX					XX
Giannella et al. (2014)	X											X
Angioli et al. (2013)	X											
Opolskiene et al. (2011)	XXX-X				XXX							
Burbos et al. (2011)	X			X								X
Burbos et al. (2010)	X			X								X
Weber et al. (1999)	XXX			X[1]X[1]	X	X	XXX		XX			

	Endometrial thickness	Hypertension	Diabetes	VAS	Vascularity	Use of warfarin	Ill-defined endometrium-myometrium interface	Irregular endometrial midline	Heterogeneous endometrium	Presence of symptoms	HE4 levels
Risk models											
Fortner et al. (2017)											
Husing et al. (2016)											
Pfeiffer et al. (2013)											
Diagnostic models											
Madkour (2017)	X				X		X	X	X		
Wong et al. (2016)	X										
Giannella et al. (2014)	X	X									
Angioli et al. (2013)	X									X	X
Opolskiene et al. (2011)	XXX			X	X	XX					
Burbos et al. (2011)			X								
Burbos et al. (2010)	X		X								
Weber et al. (1999)		X	XXX								

Abbreviations: BMI (body mass index), FFTP (first full-term pregnancy), HRT (hormone replacement therapy), OC (oral contraceptives), VAS (visual analogue scale).
[1] Weight was included as predictor instead of BMI.
[2] Biomarkers included: adiponectin, oestrone, IL1Ra, TNFα and triglycerides.

prospective cohort design. The sizes of the study populations were large: 146,679 (1559 with cancer) (Pfeiffer et al., 2013) and 201,811 (855 with cancer) (Husing et al., 2016). Inclusion criteria were similar for the two studies.

3.1.2. Outcome and predictors

Cases were identified through record linkage with regional cancer registries, linkage to health insurance records, active follow-up of study subjects and systematic requests of patient records from pathology registries in the European study (Husing et al., 2016). Identification of cases in the other study was done via linkage with state cancer registries, annual study updates and reviews of medical records (Pfeiffer et al., 2013). Both studies included predictors that were previously associated with endometrial cancer. The European study handled most of the continuous predictors as a linear term and centered them at the median (Husing et al., 2016), while the other study categorized all continuous predictors (Pfeiffer et al., 2013). Both studies investigated interaction effects between predictors and had an EPV above ten.

3.1.3. Model development

Both studies encountered incomplete data. The US study created an indicator variable for the predictor (benign breast disease) with 20% missing data and excluded all women with missing data for other predictors (Pfeiffer et al., 2013). The other study used a single, simple imputation for the missing data (Husing et al., 2016). Both prediction models were developed with Cox proportional hazard regression for the relative risks with an additional cause specific competing risk analysis to enable computation of the age specific absolute risk of developing endometrial cancer. The two studies used a stepwise backward selection procedure to identify the strongest predictors, with alpha 0.01 (Pfeiffer et al., 2013) and 0.1 (Husing et al., 2016). The developed models both included the predictors age at menopause, BMI, parity, (duration of) oral contraceptives and menopausal hormone therapy (MHT) use, and smoking. The European model also included age at menarche, age at first full term pregnancy, and an interaction term between age at menarche and BMI. The US model included an interaction term of MHT use and BMI $< 25\ kg/m^2$ (Table 2). The European study used a bootstrap sample procedure followed by linear shrinkage to improve internal validity (Husing et al., 2016).

3.1.4. Model performance and validation

The US study used independent data from the Nurses' Health Study (NHS) to assess model performance. The validation dataset consisted of 37,241 participants and 532 cases. Performance of the risk model in the external dataset was assessed with the c-statistic (0.68 [0.66–0.70]) and expected versus observed ratio (E/O ratio) (1.20 [1.11–1.30]) (Table 3).

The European study used five-fold cross validation; no specification was given about the split of the data. The discriminative ability was assessed with the c-statistic (0.77 [0.68–0.85]) and the calibration with the E/O ratio (0.99) and the Hosmer-Lemeshow test ($p = 0.08$). The integrated discrimination index (IDI) was also estimated (0.18% [0.04-

Table 3
Model performance and validation measures.

	E/O ratio	Hosmer-Lemeshow test	c-statistic	Internal validation	External validation
Risk models					
Fortner et al. (2017)			0.64 [0.60–0.69]	X	
Husing et al. (2016)	X	X	0.77 [0.68–0.85]	X	
Pfeiffer et al. (2013)	X		0.68 [0.66–0.70]		X
Diagnostic models					
Madkour (2017)			0.95		
Wong et al. (2016)			0.71 [0.66–0.75]	X	
			0.93 [0.90–0.95]	X	
Giannella et al. (2014)			0.88 [0.84–0.91]	X	
Angioli et al. (2013)			0.957 [0.91–0.98]	X	X[2]
Opolskiene et al. (2011)			0.74 [0.67–0.81]		
			0.82 [0.76–0.87]		
			0.89 [0.84–0.94]		
			0.91 [0.87–0.95]		X[1]
Burbos et al. (2011)			0.73 [0.70–0.77]		X[1]
Burbos et al. (2010)			0.77	X	
Weber et al. (1999)			0.75	X	
			0.74	X	
			0.66	X	

[1] = external validation was performed by Sladkevicius et al.
[2] = external validation was performed by Plotti et al.

0.3]) to examine the difference between the developed model and a model that only included age and country (Table 3).

The European model was extended with serum-based biomarkers in a separate study (Fortner et al., 2017). The same population (EPIC cohort) was used for updating, with a nested case-cohort design (716 participants, 247 cases). The biomarkers include adiponectin, total cholesterol, HDL cholesterol, C-peptide, C-reactive protein, androstenedione, DHEAS, oestrone, glucose, IGFBP1, IGFBP2, IL1Ra, IL6, SHBG, testosterone, TNF receptor 1, TNF receptor 2, TNFα and triglycerides. The biomarkers were log2-transformed and adjusted for age, center and menopausal status. Missing data was imputed with mean values. A backward selection procedure with alpha 0.157 was used to select serum-based biomarkers. The improvement of the model was assessed with the c-statistic and showed 0.02 points improvement (from 0.627 to 0.647, corrected for optimism) for the model with all biomarkers, and 0.017 points improvement (from 0.627 to 0.644, corrected for optimism) for the model with selected biomarkers (Table 3).

3.2. Prediction models for endometrial cancer for symptomatic women (diagnostic models)

3.2.1. Study designs and population

Seven out of eight studies developed diagnostic models in cohorts from Europe (n = 5), the Middle East (n = 1) or Hong Kong (n = 1). One study used a case-control design for model development with participants from the US (Weber et al., 1999) (Table 1). Population size varied considerably between studies, from 60 (10 cases) to 4383 (168 cases). Seven studies included women presenting with postmenopausal bleeding. Postmenopausal bleeding was in four studies defined as vaginal bleeding after at least one year of spontaneous amenorrhoea (Burbos et al., 2010, 2011; Opolskiene et al., 2011; Sladkevicius and Valentin, 2016). Two studies extended this definition with a minimum age of 40 years (Giannella et al., 2014; Madkour, 2017) and one study did not specify any definition (Wong et al., 2016). Two studies only included women with postmenopausal bleeding and endometrial thickness larger than 4 mm (Giannella et al., 2014) or 4.5 mm (Opolskiene et al., 2011). One study included women aged 40–65 years with ultrasound endometrial abnormalities (endometrial thickness, polyps and submucous myoma) and scheduled hysteroscopy (Angioli et al., 2013).

3.2.2. Outcome and predictors

The diagnosis of endometrial cancer in all studies was based on histopathology of the tissue that was obtained during endometrial sampling. Two studies restricted endometrial sampling to women with endometrial thickness over 5 mm, based on transvaginal ultrasonic scanning (Burbos et al., 2010, 2011).

All studies included age as predictor in the multivariable modelling, complemented with epidemiological predictors, such as BMI and parity, or clinical predictors such as endometrial thickness and the presence of hypertension. One study also included HE4 (Human Epididymis Protein 4) levels, a tumour marker that can be obtained via a blood test. Six out of eight studies had an EPV of more than ten, the other studies had an EPV of 8 (Weber et al., 1999) and below 1 (Madkour, 2017). Continuous predictors were kept continuous in seven studies (Wong et al., 2016; Burbos et al., 2010; Giannella et al., 2014; Weber et al., 1999; Burbos et al., 2011; Opolskiene et al., 2011; Angioli et al., 2013), one study dichotomized the continuous predictors (Madkour, 2017).

3.2.3. Model development

Five studies (Wong et al., 2016; Burbos et al., 2010; Weber et al., 1999; Burbos et al., 2011; Opolskiene et al., 2011) mentioned the presence of missing data, but the handling of missing data was described in only one study (Wong et al., 2016). Wong et al. imputed the mean for some variables and assigned the category 'multiparous' to women with missing data for the variable parity, as preliminary analysis showed no significance for this missing category. All studies used logistic regression analyses for model development. Six studies performed a preselection of predictors based on univariable analyses, with a p-value of 0.20 (Wong et al., 2016; Weber et al., 1999) or p-value of 0.05 (Giannella et al., 2014; Opolskiene et al., 2011; Madkour, 2017; Angioli et al., 2013). The methods and criteria for selection of predictors during multivariable modelling varied among the studies. Forward selection, backward selection and full model approaches were used. Selection was based on p-values of 0.05 in five studies (Wong et al., 2016; Giannella et al., 2014; Weber et al., 1999; Opolskiene et al., 2011; Angioli et al., 2013). One study did not clearly report the methods that were used during multivariable modelling (Madkour, 2017). The predictors that were included in the developed models varied, ranging from models with mostly epidemiological predictors, to models that only included predictors related to abnormalities of the endometrium (Table 2).

3.2.4. Model performance and validation

Performance of the diagnostic models was assessed with measures specific for diagnostic tests such as sensitivity and specificity (n = 7), positive and negative likelihood ratio (n = 6), positive and negative predictive value (n = 5) and the Youden index (n = 2). All studies used the area under the ROC curve (AUC), which is the same as the c-statistic in logistic regression analysis, to describe the discriminative ability of the prediction model. The AUC varied from 0.73 to 0.957. The goodness-of-fit was assessed with the Hosmer-Lemeshow test in two studiesx (Giannella et al., 2014; Weber et al., 1999). Three studies (Wong et al., 2016; Giannella et al., 2014; Angioli et al., 2013) used internal validation with a split-sample method, with a split into two equal parts (Giannella et al., 2014) or in three parts (Wong et al., 2016; Angioli et al., 2013) (two third for model development, one third for validation).

One study (Sladkevicius and Valentin, 2016) assessed the added value of endometrial thickness to a model that included only clinical predictors. Two external validation studies of diagnostic models were found. Both studies were temporal validations, as data for model development and external validation was collected in the same hospital but data used in the validation was more recent. The authors of the external validation studies were involved in the original model development. Both studies used a prospective cohort to validate the developed model with sample sizes of 80 women with cancer (Sladkevicius and Valentin, 2016) and 102 women with cancer (Plotti et al., 2017). One study described the performance of the models in the external data with the AUC, which was comparable to the performance in the original data(AUC of 0.89 and 0.91 (0.91 and 0.89 in the original data respectively)) (Sladkevicius and Valentin, 2016). The calibration was shown with calibration plots, without further specification of the intercept and slope. One study described the model performance with the predicted and observed number of malignant cases (93 predicted versus 102 observed) and benign cases (187 predicted versus 196 observed) and used a predefined cut-off point to determine the sensitivity, specificity, and the positive and negative predictive value. The model showed improved performance in terms of sensitivity (94% versus 89%) and positive predictive value (0.91 versus 0.73).

4. Discussion

This review shows the reported prediction models for risk of endometrial cancer in the general population and presence of endometrial cancer in symptomatic women. Most models were developed and validated in European and Northern American populations. One of the two risk models was externally validated in a large independent cohort (Pfeiffer et al., 2013). Only three of the fourteen diagnostic models were externally validated in a temporal validation study (Sladkevicius and Valentin, 2016; Plotti et al., 2017). No head to head comparison of developed models was found.

Age was included in almost all models; BMI and parity were also frequently included. The risk models further included variables related to the reproductive history of women, hormone use, and smoking. All these predictors are known for the relation with developing endometrial cancer (Smith et al., 2003; Beral et al., 2005; Lesko et al., 1985; Hinkula et al., 2002; Kitson et al., 2017). The relation between age and the development of several types of cancer has been most extensively described (DePinho, 2000; Balducci and Ershler, 2005; Anisimov, 2003). In the diagnostic models clinical predictors were more important, such as endometrial thickness, recurrent bleeding and insulin resistance.

Shortcomings in methodology were found in almost all studies, which is consistent with the findings from other systematic reviews on prediction models in different research areas (Damen et al., 2016; Mushkudiani et al., 2008; Collins et al., 2013; Bouwmeester et al., 2012). Information on handling of this missing data was limited. Nine (Pfeiffer et al., 2013; Wong et al., 2016; Husing et al., 2016; Burbos et al., 2010; Fortner et al., 2017; Weber et al., 1999; Burbos et al., 2011; Opolskiene et al., 2011; Sladkevicius and Valentin, 2016) out of fourteen studies mentioned the presence of missing data, of which four studies (Pfeiffer et al., 2013; Wong et al., 2016; Husing et al., 2016; Fortner et al., 2017) described the handling of the missing data. Handling varied from creating an indicator variable for the missing data to single, simple imputation. A previous simulation study has shown that the indicator method will lead to biased results, even when the missing data is missing completely at random (Donders et al., 2006). None of the models used multiple imputation, while this is considered the preferred method (Donders et al., 2006; Marshall et al., 2010).

The way of identifying the strongest predictors varied between the models. Both forward and backward procedures were used with different stopping rules. The choice for a stopping rule has an influence on the number of predictors that will be selected during the modelling process, which may result in overfitting depending on the sample size (Steyerberg, 2009). Six out of eight studies with diagnostic models (Wong et al., 2016; Giannella et al., 2014; Weber et al., 1999; Opolskiene et al., 2011; Madkour, 2017; Angioli et al., 2013) used a rather stringent p-value of 0.05 during multivariable modelling, while samples sizes were relatively small. Two of these diagnostic models had an EPV below ten, which might contribute to overfitting of the prediction model, resulting in too extreme predictions for new patients. Especially the performance (c-statistic of 0.95) of the model with EPV below 1 might be too optimistic, and will probably deteriorate during external validation.

Reproducibility of the results can be studied with internal validation (Steyerberg and Vergouwe, 2014). Bootstrapping is considered the most efficient method to study internal validity in small datasets (Moons et al., 2012). Three studies (Husing et al., 2016; Fortner et al., 2017; Weber et al., 1999) used bootstrapping to internally validate the developed models and three studies (Wong et al., 2016; Giannella et al., 2014; Angioli et al., 2013) used a split-sample. Remarkable is the use of bootstrapping in the EPIC study, as their sample size was large (855 women with cancer) and bootstrapping might therefore be considered unnecessary as overfitting and optimism is limited in such large sample sizes (Steyerberg, 2009).

It is important to assess the validity of a model in independent data from a different setting (geographical validation) or from a more recent time period (temporal validation) (Steyerberg, 2009; Altman and Royston, 2000; Justice et al., 1999). Three studies performed an external validation, in which the models showed relatively good performance in the independent data (Pfeiffer et al., 2013; Sladkevicius and Valentin, 2016; Plotti et al., 2017). One model (Pfeiffer et al., 2013) overestimated the number of cases, indicating suboptimal calibration in the large. This was related to the difference in average risk in the population that was used for model development. Miscalibration in the large is often found and can easily be adjusted for (te Velde et al., 2014; Vergouwe et al., 2010).

The two risk models included in this review have moderate discriminative ability, which did not improve substantially by updating the model with serum-based biomarkers. This implies that new information may be needed to improve the model performance. Improvement is necessary because implementing a model with poor discriminative ability has little value in practice. New information can for example be found in the field of epigenetics. Current developments in other research areas have already shown that the epigenome contains objective information on environmental exposure and can be useful for making risk predictions. Alterations in DNA methylation were shown to be the consequence of adiposity. The change in DNA methylation predicted the risk of developing type 2 diabetes (Wahl et al., 2017). In addition, the hypomethylation of the aryl hydrocarbon receptor repressor (AHRR) gene holds information on former smoking status, which might contribute to risk predictions of lung cancer (Bojesen et al., 2017). Further, the methylation of peripheral blood cell DNA can serve as a predictor for the risk of developing breast cancer

(Widschwendter et al., 2008). Adding information from the epigenome to existing risk prediction models likely improves their performance, as information from the epigenome is less error-prone and may therefore hold more information than data from questionnaires (Ladd-Acosta and Fallin, 2016; Pashayan et al., 2016). Adding genetic information for instance on mutations in mismatch repair genes, as seen in Lynch syndrome, may also improve model performance. Women with Lynch syndrome have approximately a 20–60% cumulative lifetime risk of developing endometrial cancer and Lynch syndrome is responsible for 2-5% of all endometrial cancer cases (Meyer et al., 2009).

The total number of studies included in this review is relatively small, especially the number of models that predict the risk of developing endometrial cancer in the general population. Despite extensive searches no additional studies were found. The small number of studies may have an influence on the overview of selected predictors, as more studies might give a more complete and representative reflection of important predictors.

In conclusion, only a few models have been developed to predict endometrial cancer in asymptomatic or symptomatic women. The usefulness of most of the models is unclear considering methodological shortcomings and lack of external validation and head to head comparisons of models. Developed risk models should be externally validated and extended with new predictors, such as genetic and epigenetic risk predictors, to improve model performance. Future research on diagnostic models should focus on external validation and creating models with larger sample sizes, which could be realized with individual patients data meta-analysis.

S1 – Search strategy

('endometrium tumor'/exp OR (((endometri*) NEAR/3 (cancer* OR neoplas* OR tumo* OR malign* OR carcinom* OR adenocarcinom* OR sarcoma*))):ab,ti**) AND ((**"Risk function" OR 'risk assessment'/exp OR "risk functions" OR "risk equation*" OR "risk chart*" OR (risk NEAR/3 tool*) OR "risk assessment function*" OR "risk assessor" OR "risk calculation*" OR "risk calculator*" OR "risk factor* calculator*" OR "risk factor* calculation*" OR "risk table*" OR "risk threshold*" OR "risk scoring method*" OR "scoring scheme*" OR "risk scoring system*" OR "risk prediction*" OR "predictive instrument*" OR "project* risk*"**) OR ((**'decision support system'/exp OR 'algorithm'/exp OR algorithm* OR algorythm* OR "predictive model*" OR "prognos* model*" OR "treatment decision*" OR "scoring method*" OR (prediction* NEAR/3 method*) OR (prognos* OR incidence* NEAR/3 model*)**) AND (**'risk factor'/exp OR 'risk assessment'/exp OR (risk* NEAR/1 assess*) OR "risk factor*"**))) NOT (**[Conference Abstract]/lim OR [Letter]/lim OR [Note]/lim OR [Editorial]/lim).

Financial support

The project leading to this application has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 634570 (FORECEE).

Conflict of interest statement

Maaike Alblas, Ewout W. Steyerberg and Yvonne Vergouwe report grants from European Union's Horizon 2020 Research and Innovation Programme during the conduct of the study; Dr. Widschwendter reports grants from University College London, during the conduct of the study; Kimberley B. Velt and Nora Pashayan have nothing to disclose.

References

Altman, D.G., Royston, P., 2000. What do we mean by validating a prognostic model? Stat. Med. 19, 453–473.
Angioli, R., Capriglione, S., Aloisi, A., Luvero, D., Cafa, E.V., Dugo, N., et al., 2013. REM (risk of endometrial malignancy): a proposal for a new scoring system to evaluate risk of endometrial malignancy. Clin. Cancer Res. 19, 5733–5739.
Anisimov, V.N., 2003. The relationship between aging and carcinogenesis: a critical appraisal. Crit. Rev. Oncol. Hematol. 45, 277–304.
Balducci, L., Ershler, W.B., 2005. Cancer and ageing: a nexus at several levels. Nat. Rev. Cancer 5, 655–662.
Beral, V., Bull, D., Reeves, G., Million Women Study C, 2005. Endometrial cancer and hormone-replacement therapy in the Million Women Study. Lancet 365, 1543–1551.
Bojesen, S.E., Timpson, N., Relton, C., Davey Smith, G., Nordestgaard, B.G., 2017. AHRR (cg05575921) hypomethylation marks smoking behaviour, morbidity and mortality. Thorax 72, 646–653.
Bouwmeester, W., Zuithoff, N.P., Mallett, S., Geerlings, M.I., Vergouwe, Y., Steyerberg, E.W., et al., 2012. Reporting and methods in clinical prediction research: a systematic review. PLoS Med. 9, 1–12.
Burbos, N., Musonda, P., Giarenis, I., Shiner, A.M., Giamougiannis, P., Morris, E.P., et al., 2010. Predicting the risk of endometrial cancer in postmenopausal women presenting with vaginal bleeding: the Norwich DEFAB risk assessment tool. Br. J. Cancer 102, 1201–1206.
Burbos, N., Musonda, P., Duncan, T.J., Crocker, S.G., Morris, E.P., Nieto, J.J., 2011. Estimating the risk of endometrial cancer in symptomatic postmenopausal women: a novel clinical prediction model based on patients' characteristics. Int. J. Gynecol. Cancer 21, 500–506.
Collaboration NCDRF, 2016. Trends in adult body-mass index in 200 countries from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 19.2 million participants. Lancet 387, 1377–1396.
Collins, G.S., Omar, O., Shanyinde, M., Yu, L.M., 2013. A systematic review finds prediction models for chronic kidney disease were poorly reported and often developed using inappropriate methods. J. Clin. Epidemiol. 66, 268–277.
Damen, J.A., Hooft, L., Schuit, E., Debray, T.P., Collins, G.S., Tzoulaki, I., et al., 2016. Prediction models for cardiovascular disease risk in the general population: systematic review. Bmj 353, i2416.
DePinho, R.A., 2000. The age of cancer. Nature 408, 248–254.
Dixon, J.B., 2010. The effect of obesity on health outcomes. Mol. Cell Endocrinol. 316, 104–108.
Donders, A.R., van der Heijden, G.J., Stijnen, T., Moons, K.G., 2006. Review: a gentle introduction to imputation of missing values. J. Clin. Epidemiol. 59, 1087–1091.
Ferlay, J., Steliarova-Foucher, E., Lortet-Tieulent, J., Rosso, S., Coebergh, J.W., Comber, H., et al., 2013. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. Eur. J. Cancer 49, 1374–1403.
Fortner, R.T., Husing, A., Kuhn, T., Konar, M., Overvad, K., Tjonneland, A., et al., 2017. Endometrial cancer risk prediction including serum-based biomarkers: results from the EPIC cohort. Int. J. Cancer 140, 1317–1323.
Giannella, L., Mfuta, K., Setti, T., Cerami, L.B., Bergamini, E., Boselli, F., 2014. A risk-scoring model for the prediction of endometrial cancer among symptomatic postmenopausal women with endometrial thickness & 4 mm. Biomed. Res. Int. 2014, 130569.
Gull, B., Karlsson, B., Milsom, I., Granberg, S., 2003. Can ultrasound replace dilation and curettage? A longitudinal evaluation of postmenopausal bleeding and transvaginal sonographic measurement of the endometrium as predictors of endometrial cancer. Am. J. Obstet. Gynecol. 188, 401–408.
Hecht, J.L., Mutter, G.L., 2006. Molecular and pathologic aspects of endometrial carcinogenesis. J. Clin. Oncol. 24, 4783–4791.
Hinkula, M., Pukkala, E., Kyyronen, P., Kauppila, A., 2002. Grand multiparity and incidence of endometrial cancer: a population-based study in Finland. Int. J. Cancer 98, 912–915.
Husing, A., Dossus, L., Ferrari, P., Tjonneland, A., Hansen, L., Fagherazzi, G., et al., 2016. An epidemiological model for prediction of endometrial cancer risk in Europe. Eur. J. Epidemiol. 31, 51–60.
Ingui, B.J., Rogers, M.A.M., 2001. Searching for clinical prediction rules in medline. J. Am. Med. Inf. Assoc.: JAMIA 8, 391–397.
Justice, A.C., Covinsky, K.E., Berlin, J.A., 1999. Assessing the generalizability of prognostic information. Ann. Intern. Med. 130, 515–524.
Kitson, S.J., Evans, D.G., Crosbie, E.J., 2017. Identifying high-risk women for endometrial cancer prevention strategies: proposal of an endometrial cancer risk prediction model. Cancer Prev. Res. (Phila.) 10, 1–13.
Kyrgiou, M., Kalliala, I., Markozannes, G., Gunter, M.J., Paraskevaidis, E., Gabra, H., et al., 2017. Adiposity and cancer at major anatomical sites: umbrella review of the literature. Bmj 356, j477.
Ladd-Acosta, C., Fallin, M.D., 2016. The role of epigenetics in genetic and environmental epidemiology. Epigenomics 8, 271–283.
Lesko, S.M., Rosenberg, L., Kaufman, D.W., Helmrich, S.P., Miller, D.R., Strom, B., et al., 1985. Cigarette smoking and the risk of endometrial cancer. N. Engl. J. Med. 313, 593–596.
MacMahon, B., 1974. Risk factors for endometrial cancer. Gynecol. Oncol. 2, 122–129.
Madkour, N.M., 2017. An ultrasound risk-scoring model for prediction of endometrial cancer in post-menopausal women (using IETA terminology). Middle East. Fertil. Soc. J. 22, 201–205.
Marshall, A., Altman, D.G., Royston, P., Holder, R.L., 2010. Comparison of techniques for handling missing covariate data within prognostic modelling studies: a simulation study. BMC Med. Res. Methodol. 10, 7.
Meads, C., Ahmed, I., Riley, R.D., 2012. A systematic review of breast cancer incidence risk prediction models with meta-analysis of their performance. Breast Cancer Res. Treat. 132, 365–377.
Meyer, L.A., Broaddus, R.R., Lu, K.H., 2009. Endometrial cancer and Lynch syndrome: clinical and pathologic considerations. Cancer Control 16, 14–22.
Moons, K.G., Kengne, A.P., Woodward, M., Royston, P., Vergouwe, Y., Altman, D.G.,

et al., 2012. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. Heart 98, 683–690.
Mushkudiani, N.A., Hukkelhoven, C.W., Hernandez, A.V., Murray, G.D., Choi, S.C., Maas, A.I., et al., 2008. A systematic review finds methodological improvements necessary for prognostic models in determining traumatic brain injury outcomes. J. Clin. Epidemiol. 61, 331–343.
Ng, M., Fleming, T., Robinson, M., Thomson, B., Graetz, N., Margono, C., et al., 2014. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 384, 766–781.
Opolskiene, G., Sladkevicius, P., Valentin, L., 2011. Prediction of endometrial malignancy in women with postmenopausal bleeding and sonographic endometrial thickness > / = 4.5 mm. Ultrasound Obstet. Gynecol. 37, 232–240.
Pashayan, N., Reisel, D., Widschwendter, M., 2016. Integration of genetic and epigenetic markers for risk stratification: opportunities and challenges. PerMed 13, 93–95.
Pfeiffer, R.M., Park, Y., Kreimer, A.R., Lacey Jr., J.V., Pee, D., Greenlee, R.T., et al., 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, e1001492.
Plotti, F., Capriglione, S., Terranova, C., Montera, R., Scaletta, G., Lopez, S., et al., 2017. Validation of REM score to predict endometrial cancer in patients with ultrasound endometrial abnormalities: results of a new independent dataset. Med. Oncol. 34, 82.
Sladkevicius, P., Valentin, L., 2016. Prospective validation of two mathematical models to calculate the risk of endometrial malignancy in patients with postmenopausal bleeding and sonographic endometrial thickness > /=4.5 mm. Eur. J. Cancer 59, 179–188.
Smit, H.A., Pinart, M., Anto, J.M., Keil, T., Bousquet, J., Carlsen, K.H., et al., 2015. Childhood asthma prediction models: a systematic review. Lancet Respir. Med. 3, 973–984.
Smith, J.S., Green, J., Berrington de Gonzalez, A., Appleby, P., Peto, J., Plummer, M., et al., 2003. Cervical cancer and use of hormonal contraceptives: a systematic review. Lancet 361, 1159–1167.
Steyerberg, E.W., 2009. Clinical Prediction Models. Springer-Verlag, New York.
Steyerberg, E.W., Vergouwe, Y., 2014. Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur. Heart J. 35, 1925–1931.
te Velde, E.R., Nieboer, D., Lintsen, A.M., Braat, D.D., Eijkemans, M.J., Habbema, J.D., et al., 2014. Comparison of two models predicting IVF success; the effect of time trends on model performance. Hum. Reprod. 29, 57–64.
Vergouwe, Y., Moons, K.G., Steyerberg, E.W., 2010. External validity of risk models: use of benchmark values to disentangle a case-mix effect from incorrect coefficients. Am. J. Epidemiol. 172, 971–980.
Wahl, S., Drong, A., Lehne, B., Loh, M., Scott, W.R., Kunze, S., et al., 2017. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. Nature 541, 81–86.
Weber, A.M., Belinson, J.L., Piedmonte, M.R., 1999. Risk factors for endometrial hyperplasia and cancer among women with abnormal bleeding. Obstet. Gynecol. 93, 594–598.
Widschwendter, M., Apostolidou, S., Raum, E., Rothenbacher, D., Fiegl, H., Menon, U., et al., 2008. Epigenotyping in peripheral blood cell DNA and breast cancer risk: a proof of principle study. PLoS One 3, e2656.
Wong, A.S., Cheung, C.W., Fung, L.W., Lao, T.T., Mol, B.W., Sahota, D.S., 2016. Development and validation of prediction models for endometrial cancer in postmenopausal bleeding. Eur. J. Obstet. Gynecol. Reprod. Biol. 203, 220–224.