TÜRK JİNEKOLOJİK ONKOLOJİ DERGİSİ
2023-1, Sayfa 9-24
Araştırma

Endometrium Kanserinde Bireyselleştirilmiş Risk Tahmin Modelinin Karar Eğrisi Analizinde Klinik Kullanılabilirliğinin Değerlendirilmesi

Assessment of Clinical Utility in Decision Curve Analysis for an Individualized Risk Prediction Model of Endometrial Cancer

Oğuz Kaan Köksal[1], Evrim Erdemoğlu[1], Volkan Öztürk[1], Kemal Kürşat Bozkurt[2], İlyas Turan[1]

ÖZET

Giriş: Anormal uterin kanaması olan kadınlar için bireyselleştirilmiş bir risk değerlendirme yöntemine (IRPM) ihtiyaç vardır. Anormal vajinal kanamanın yönetimi, pozitif/negatif prediktif değerlere sahip tanısal test sonuçlarına dayanır. Bunun hastada biyopsi için harekete geçip geçmeme eşiğinde katkısı yoktur. Bu çalışmanın amacı, hastanın karar sürecine katkıda bulunabilecek, ayırt etme yeteneğine sahip kalibre edilmiş bir model geliştirmektir.

Gereç ve Yöntem: Bu çalışma 35 yaşından büyük hastaların verilerini çıkarmak için kesitsel one gate bir kohort çalışması olarak planlanmıştır. Tüm hastalara indeks testi (endometriyal kalınlık ölçümü) ve referans testi (genel anestezi altında D&C) uygulanmış, WHO 2014'e göre bening veya premalign olarak histopatolojik raporlar ayrılmıştır. Birincil amaç, yararlı bir klinik IRPM geliştirmek ve IRPM'nin net faydasını ve karar eğrisi analizinde çeşitli hastalık eşikleri için mevcut kılavuzları taklit eden modelleri karşılaştırmaktır.

İkincil amaç ise, ayırt edici özellikleri, gereksiz biyopsilerin sayısını ve çeşitli eşiklerde kaçırılan vakaları analiz etmektir.

Bulgular: Semptom durumu, endometriyal kalınlık ve yaştan oluşan IRPM, pre-/endometrium kanseri için en iyi risk öngörme yöntemi olarak bulunmuştur. IRPM, klinik eşik olasılıklarının tüm aralığından biyopsi alma veya almama konusunda yönergelerden daha yüksek bir net faydaya sahiptir. IRPM gözden kaçan vakaların sayısını da azaltabilir. Mevcut yönergeleri taklit eden modellerin yalnızca %3'lük bir eşiğin üzerinde faydalı olduğu ve bu eşiğin altında zararlı olabileceği bulunmuştur. IRPM herhangi bir ek maliyete veya zaman alıcı analize ihtiyaç duymamaktadır.

Sonuç: IRPM herhangi bir ek maliyete veya zaman alıcı analize ihtiyaç duymaz. IRPM, hastanın daha ileri inceleme kararına katkıda bulunmasına yardımcı olabilir. IRPM'nin klinik kullanışlılığı, mevcut uygulamayı taklit eden modellerden daha üstündür. Heterojen bir hasta grubunda kılavuzların tanısal ayrımcı kesme değerleri beklenenden daha düşüktür.

Anahtar sözcükler: Karar Eğrisi Analizi, Endometrium Kanseri, Bireyselleştirilmiş Risk Tahmin Modeli

ABSTRACT

Background: There is a need for a personalized risk assessment method (IRPM) for women with abnormal uterine bleeding. Management of abnormal vaginal bleeding is based on diagnostic test results with positive/negative predictive values for a cut-off value. Patient has no contribution on the threshold to act for biopsy or not. Aim of present study was to develop a calibrated model with discriminative ability in which the patient can contribute to decision process.

Methods: A cross-sectional one-gate design cohort study was planned to extract data of patients older than 35 years-old. All had index test (endometrial thickness measurement) and reference test (D&C under general anesthesia). Target was histopathological report according to WHO 2014 as benign or pre/cancer. Primary outcome was to develop a useful clinical IRPM and to compare net benefit of IRPM and models mimicking current guidelines for various thresholds of disease in decision curve analysis. Secondary outcomes were to analyse discriminative properties, the number of unnecessary biopsies and missed cases at various thresholds.

Findings: IRPM consisting of symptom status, endometrial thickness and age was the best risk predicting method for pre-/endometrium cancer. IRPM had a higher net benefit than guidelines and to biopsy all or not at entire range of clinical threshold probabilities. IRPM had a good discrimination slope and can also decrease number of missed cases. Models mimicking current guidelines were only useful above a threshold of 3% and below this threshold, it is found to be harmful.

Interpretation: IRPM doesn't need any additional costs or time-consuming analysis. IRPM may aid patient to contribute to decision of further investigation. Clinical usefulness of IRPM is superior to models mimicking current practice. Value of diagnostic discriminatory cut-offs of guidelines is lower than expected in a heterogenous group of patients.

Key words: Endometrial Cancer, Risk Prediction, Decision Curve Analysis

Geliş Tarihi: January 24, 2023 Kabul Tarihi: April 15, 2023

[1]Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, Süleyman Demirel University, Isparta, Turkey
[2]Department of Pathology, Süleyman Demirel University, Isparta, Turkey

İletişim: Evrim Erdemoglu
Division of Gynecologic Oncology, Süleyman Demirel University, Isparta, Turkey

Tel: +90 505 272 4344
E-posta: evrimmd@yahoo.com
ORCID ID: 0000-0002-5440-2940

Research in Context

Management of women with abnormal bleeding is based on diagnostic test results with positive and negative predictive values for cut-off value and the patient has no contribution on the threshold to act for biopsy or not. However, there is a need for a personalized risk assessment method for women with abnormal uterine bleeding, as well as to compare clinical utility and risk prediction success of different models.

A risk prediction model gives predicted probability directly and the clinical usefulness for different thresholds can be evaluated. Besides, multivariable risk prediction tools can significantly reduce unnecessary systematic biopsies, without compromising the detection of pre/cancer.

Evidence Before This Study

Previous reports are based on discriminative properties and these studies have various unstandardized methodologies with major drawbacks. Risk-based patient selection for biopsy in abnormal bleeding has been adopted in daily clinical practice by clinical judgment and endometrial thickness (ET) measurement via transvaginal ultrasonography. Clinicians refer to cut-off values suggested by guidelines that mainly cover postmenopausal women or literature-based results to decide if an endometrial biopsy should be undertaken. Although advancement in the technology, incidence and mortality of endometrial cancer is increasing.

A risk prediction model developed only for postmenopausal women with bleeding would not be ideal for everyday practice and such a design would miss symptomatic premenopausal and asymptomatic postmenopausal patients. Nearly 25-40% of endometrial cancer and the majority of EIN cases are seen in premenopausal women. Secondly, categorizing patients by menopausal status (premenoapusal and postmenopausal) and symptom status (asymptomatic/symptomatic) or dividing patients into younger or older than 45 years old assigns all patients into the same group. Besides it attributes the same arbitrary predicted risk to the patient whereas there is a wide variety of risk spectrum in each group. The current practice therefore is far away to provide a personalised risk and does not include the patient in the decision process.

Added Value of This Study

A simple model which is not costly and time consuming that can be used in the clinical practice is to use symptom status, age and endometrial thickness. This model provides an individualized risk to the patient. Clinical usefulness of this IRPM is superior to the models mimicking current practice. We were able to compare the individualized risk models' superiority to other models using abnormal bleeding, menopause and endometrial thickness >4mm. Besides, its discriminative properties are good enough to use it in daily practice. The comparator models mimicking current guidelines (Figure 2) were only useful above a threshold of 3% and below this threshold, it is found to be harmful to use the comparator model anyway. IRPM helps also to decrease missed cases and unnecessary biopsies for different risk thresholds, as well as in patients with important co-morbidities, this risk-based system allows making a comparison if the risk of cancer outweighs the risk of co-morbidities concerning the procedure.

Implications of all the available evidence

The value of diagnostic discriminatory cut-offs, suggested by guidelines is lower than expected in a heterogenous group. The IRPM can be updated and calibrated for different populations in various geographical locations, this approach is cumbersome for models based on cut-off values. IRPM does not need any additional costs or time-consuming analysis and may aid the patient to contribute to the decision of further investigation.

Introduction

Endometrial cancer is the most common gynecologic malignancy with a rising incidence in the last 20 years. The total rate of increase has reached 12% (1). Despite advancements in diagnostic technologies and treatment modalities, mortality rates due to endometrial cancer increased 21% in the last two decades (1). Women with endometrial cancer usually present with vaginal bleeding, and endometrial biopsy in abnormal bleeding has been adopted in daily clinical practice by menopausal status and endometrial thickness (ET) measurement via transvaginal ultrasonography (TVS) in order to identify women with endometrial intraepitelial neoplasia (EIN) or endometrial cancer. Clinicians refer to cut-off values of ET suggested by guidelines to decide if a biopsy should be undertaken. There are some shortcomings of this approach; biopsy is recommended to all women with postmenopausal bleeding and an ET higher than 3-5 mm. Some guidelines either have no recommendation for premenopausal women or advise biopsy to all symptomatic premenopausal women older 42 years old. However, nearly 25-40% of endometrial cancer and the majority of EIN cases are seen in premenopausal women (2). This simplified dichotomous categorization assigns the patient to a generalized group and designa-

tes an equivalant risk to all women in the same group; a 45-year-old women with postmenopausal bleeding and ET of 5-6mm has the same risk as an 80 year old women with and endometrial thickness of 20 mm.

Another important drawback of this approach is that the equivalant risk assigned to the group is arbitrary such as; a risk threshold of 3% was set for adult cancers in NICE guidelines (3). This arbitrary predetermined risk can be high for some patients, whereas it can be low or can be disregarded in some patients because of other health conditions. In the current practice neither the patient can get involved in the decision process nor the gynecologist can provide an individualized risk for the patient.

Another shortcoming is that there is not a well-designed randomized controlled trial in women with abnormal bleeding. Meta-analyses evaluating TVS in women with abnormal bleeding are based on observational studies. Observational studies have likely unidentified sources of confounding and risk modification. The meta-analyses concluding non-randomized, non-blinded studies can overemphasize the effect (4). Secondly, in these meta-analyses patients were managed due to existing guidelines. Besides, in most of the studies TVS accuracy were not verified by biopsy in all patients (5). Therefore, meta-analyses of TVS are subject to debate over the validity of meta-analytical approaches (4) and pooling such findings may not lead to more certain outcomes for diagnostic and discriminative purposes (4,5).

There is a need for a personalized risk assessment method for women with abnormal uterine bleeding. A risk prediction model gives predicted probability directly and the clinical usefulness for different thresholds can be evaluated. Besides, multivariable risk prediction tools can significantly reduce unnecessary systematic biopsies, without compromising the detection of pre/cancer. Aim of the present study to develop a calibrated model for reliable diagnostic risk prediction and discriminative ability in which the patient and doctor can contribute to the decision process.

Material and Methods

A cross-sectional one-gate design retrospective cohort study was planned to extract data between 01.01.2015 and 01.09.2020 in division of Gynecologic Oncology, department of Obstetrics and Gynecology. The study design was per Cochrane diagnostic test accuracy tools and Quads-2. The design and results were submitted in comparison with Standards for Reporting Accuracy Studies (STARD) (6) and guidelines for reporting cross-sectional studies (STROBE) (7). The study was registered and approved by IRB (Number 20/276).

Participants

Patients older than 35 years old were included in the study who had the index test (endometrial thickness measurement by TVS) and reference test (endometrial sampling) performed by one gynecologist (Dr E.E.). Target, histopathologic evaluation, was done by one gynecologic pathologist (Dr K.K.B).

Abnormal bleeding was defined by heavy menstrual bleeding, inter-menstrual bleeding, or menometrorrhagia, irregular menses, and postmenopausal bleeding. Women with abnormal bleeding were categorized as symptomatic. Patients included in the study were assigned 0 either for premenopausal or asymptomatic and were assigned 1 either for postmenopausal or symptomatic status. Menopause is defined as the absence of a menstrual cycle after 12 months from the last menstrual period.

Patients who had Lynch syndrome or followed up for a previously diagnosed endometrial pathology, had a history of fertility-preserving treatment for endometrial cancer, receiving hormone replacement therapy or selective estrogen receptor modulator, younger than 35 years old were excluded. Patients whose TVS could not be performed or who had an insufficient endometrial biopsy result were also excluded from the study. If the evaluating pathologist was different or if the index test and reference test were not performed by the same operator, the data were not recorded. The flowchart (Figure 1) shows the patients included in the study.

Index Test

TVS was performed by an experienced gynecologic oncologist. Mid-saggital section of the uterus was visualized to measure the endometrial thickness. Fluid in the endometrial cavity was measured separately and it was not included in the sum of endometrial thickness.

Predictive Factors

Age, endometrial thickness measured by transvaginal ultrasonography, abnormal uterine bleeding symptoms, history of diabetes mellitus, hypertension and smoking were abstracted. Body mass index of patients was recorded.

Reference Test

All postmenopausal patients with bleeding were undertaken to biopsy regardless of endometrial thickness. All postmenopausal asymptomatic patients with endometrial thickness ≥3mm had the reference test. Premenopausal women with abnormal uterine bleeding and

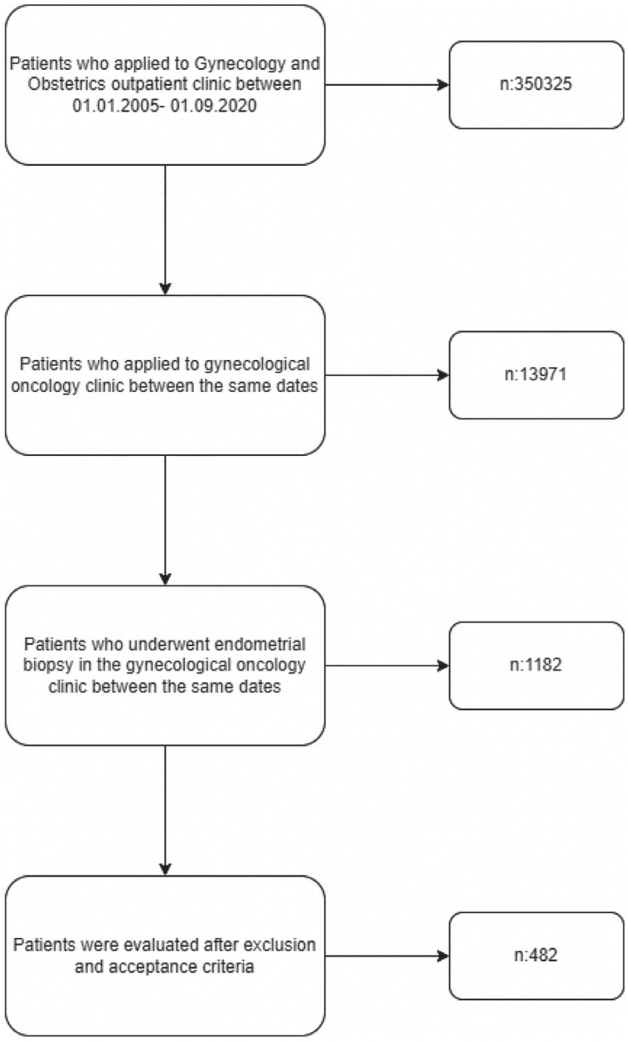


Figure 1: Flowchart of patients included to the study

premenopausal asymptomatic women with a suspected endometrial pathology or as a preoperative screening before the gynecologic operation had the reference test.

Dilatation and curettage under general anesthesia was the reference test. Cervical dilatation to at least number eight Hegar dilators was performed and then sharp curettage was performed with a medium-sized Sims curette. High vacuum curettage was undertaken by Karmann in women whose cervix could not be dilated.

Target

Histopathologic Evalution

Neutral buffered 10% formalin solution was used for the fixation of the endometrial biopsy specimens. The entire specimen was sampled for each patient. After an overnight tissue processing program paraffin blocks were prepared, and 5 µm thick sections were obtained from these paraffin blocks for the histopathological evaluation. Sections were stained with H&E and evaluated under a light microscope by a gynecologic pathologist.

Pathologic Classification

Endometrial precursor lesions were classified as follows:

Endometrial hyperplasia without atypia; decreased stroma between the crowded, branched and dilated glands, which are lined by a pseudostratified epithelium and lack cellular atypia (8).

Endometrial atypical hyperplasia/endometrioid intraepithelial neoplasia (EIN); composed of crowded and altered glands, which are cytologically different from the adjacent endometrial glands. In the EIN area glands exceed the stroma (>1/1) and the total diameter of the area must be of sufficient size (greater than 1 mm) (9).

Final report of the lesions was made by WHO 2014 classification as follow; benign lesions and hyperplasia without atypia were coded benign and atypical endometrial hyperplasia, endometrial intraepiethelial neoplasia and carcinoma as pre/cancer (10).

Outcomes

Primary outcome was to analyze independent predictors of pre-/endometrium cancer and then develop a simple useful clinical diagnostic predictor model where the patient can also be involved in the decision process of whether to take a biopsy or not. We wanted to compare the clinical utility of the developed model with comparator models mimicking current guidelines and practice. The final primary outcome was the net benefit of each model for various thresholds of disease.

The secondary outcome was to investigate the diagnostic discriminative ability of developed model and comparator models with ROC curve analysis and regression analyses. We also wanted to analyze the contribution of each predictor in different models. Thirdly, we investigated the diagnostic test accuracy parameters and pretest and posttest probabilities in confusion matrix analyses of the developed model.

Statistical Analyses

Categorical variables were summarized as counts and percentages, continuous variables were summarized as mean and standard deviation or median and range. Alpha level was set to 0.005 for all statistical analyses and a two-sided p-value less than 0.005 was considered as an important finding. Pathology reports were dichotomized to benign and pre-/cancer for regression and diagnostic analyses. All analyses, model construction and data preparation were undertaken with Wizard 2 and Datagraph. Syntax was written for bootstrapping in SPSS. Confusion matrix and diagnostic test results. with pretest and posttest probability were processed in the online tool from University of Illinois, Chicago (available at http://araw.mede.uic.edu/cgi-bin/testcalc.pl)

Correlation and Regression Analysis

For the analysis of primary outcome; Pearson correlation analysis for continuous data and rank correlation for categorical data were performed. All risk modifying factors for endometrial cancer recorded in the study and others with a p-value smaller than 0.005 in correlation test were entered in univariate and then in the multivariable logistic regression model. We preferred a discriminative classifier - logistic regression rather than naive Bayes because it returns well-calibrated predictions by default. Age, endometrial thickness and body mass index were standardized for regression analyses by subtracting the mean of each variable from the observed value and divided by standard deviation. Multicollinearity and confounders were analyzed. Separate models for multi-collinear variables were built and studied. When the odds ratio changed by 10% or more upon including a confounder in the model, the confounder was controlled by leaving it in the model.

Model Specifying

Main effects and interactions were analyzed. Model specifying was based on p values of independent variables. All candidate variables were entered and then repeatedly each term with the highest non-significant p-value were removed until the model was reduced to contain only significant variables.

In addition to our analyses described above, we have also tested the regression model with an automated selection procedure using stepwise likelihood ratios based on the significance of the score statistic, and removal testing based on the probability of a likelihood-ratio statistic based on the maximum partial likelihood estimates. The final model found in the automated selection was the same and it was named the individualized risk predictor model (IRPM - Model A).

Standardized coefficients, standard errors, odds ratios with confidence interval for each factor and p values were calculated. Diagnostic risk predictions and 95% confidence intervals were calculated and given in a user-friendly table, where 3% risk was colored differently to create a visual summary. 3% risk was chosen because a risk threshold of 3% was set for adult cancers in NICE guidelines (3).

Calibration and Validation of Developed Individualized Risk Predictor Model

Internal validation was tested by bootstrapping. Regression coefficients and confidence intervals via 10000 bootstrap resample were investigated to analyze the confidence intervals. The average predicted risk was compared with the overall event rate to assess the calibration of the model in the large. Brier score was calculated to test model performance.

Comparator Models Mimicking Current Guidelines

For evaluation of clinical utility in decision curve analysis, two models based on the current literature (11, 12) were created. These two models were mimicking the management of abnormal uterine bleeding due to guidelines and literature. The flowchart and attributed

risks of patients in these models are shown in figure 2. Patients were categorized as pre or postmenopausal and then stratified to find their attributed risk (model B- categoric) or to find if they fall into biopsy zone or no biopsy zone (model C- categoric).

In categoric model B, patients that fall into no biopsy zone are given a risk of <1% pre/cancer risk. Risk percentages of women that fall into the biopsy zone were; 2.6% for premenopausal women with abnormal bleeding and ≥45 years old, 4.6% for postmenopausal bleeding and endometrial thickness >4mm, 6.7% for postmenopausal asymptomatic women with an endometrial thickness >11mm (11, 12).

In categoric model C, no biopsy was suggested to patients that fall into no biopsy zone, and biopsy was suggested to all patients that fall into the biopsy zone.

A comparator model for c-statistics was created by setting the omitted variables of independent predictors to no abnormal bleeding, premenopausal and endometrial thickness ≤4mm (AUROC comparator model).

Statistical Evaluation of Models for Diagnosis; Discrimination

Discriminative ability was assessed by c-statistics, discrimination slope, violin plots and whisker box plot of predicted probabilities. Area under the receiver-operating characteristic curve (AUROC) metrics were produced and change of AUC by addition of each predictor was calculated for the developed model and comparator model.

Clinical Utility of Models and Comparison

Diagnostic and prognostic models are typically evaluated with measures of accuracy that do not address clinical consequences. The appropriate thresholds are different for each policy and guideline and can differ for each patient. We have used net benefit and decision curve analysis. Net benefit is a tool for evaluating the clinical implications of models, markers, and tests. Current management and guidelines do not take into account patients preferences. Women should generally be involved in decision making about their care, such as whether or not to have a biopsy for cancer. A key concept in DCA is that of a probability threshold, namely, a level of diagnostic certainty above which the patient would choose to be treated. DCA also allows the comparison of different models for various thresholds.

Net benefit= *TruePositives/n–Falsepositives/*(*pt*1,*pt*)

We have compared the individualized model (Model A) with categoric model B, model C and taking biopsy all patients or none of the patients and plotted it on decision curve analysis.

Role Of the Funding Source

All authors had full access to all of the data and the final responsibility to submit for publication and had no conflict of interest.

Results

Primary Outcome 1.1; Development of Model

Demographic characteristics of patients are shown in Table 1. Age, endometrial thickness, BMI, hypertension, history of abnormal bleeding and menopausal status, smoking were correlated with target results. Univariate analyses of these predictors are shown in figure 3 and

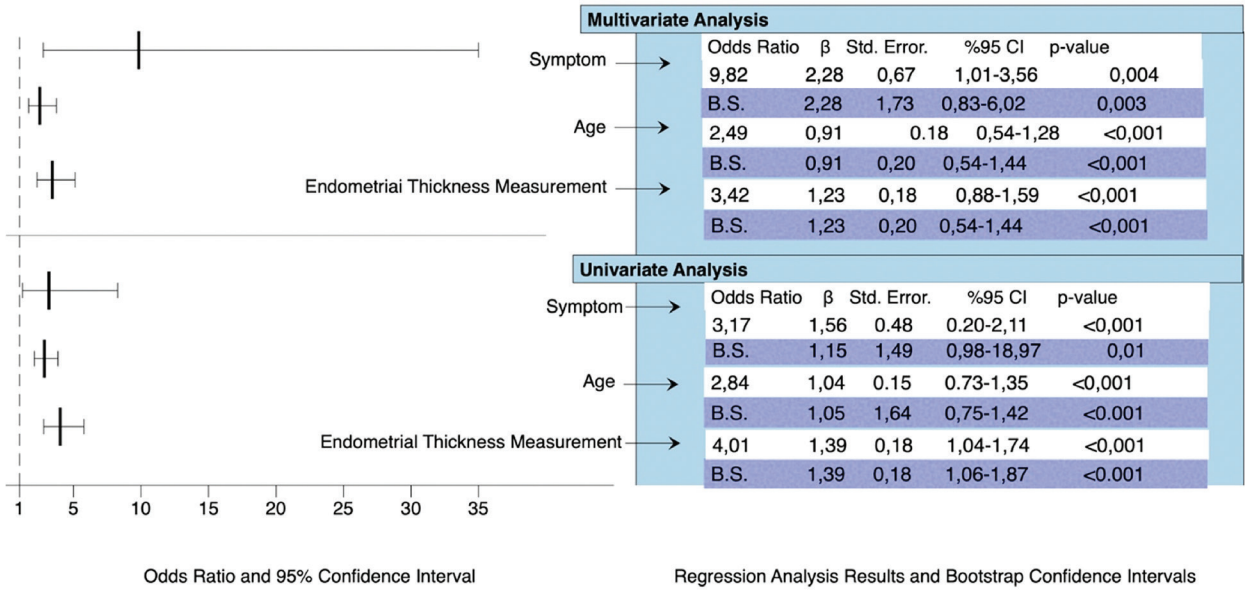


Figure 2: Management of patients by categoric model B and categoric model C

Table 1 • Charectersitics of Patients.

	Perimenopause (n:297)		Postmenopause(n:185)	
	Asymptomatic (n:72)	**Symptomatic (n:225)**	**Asymptomatic (n:68)**	**Symptomatic (n:117)**
Age (median)	**44 (±5)**	**45 (±5)**	**62 (±9)**	**59 (±10)**
ET	10 (±5)	8 (±4)	10 (±6)	10 (±10)
HT	8 (%11,11)	26 (%11,56)	31 (%45,59	50 (%42,74)
DM	8 (%11,11)	19 (%8,44)	23 (%33,82)	25 (%21,37)
Smoking	11 (%15,28)	45 (%20)	3 (%4,41)	9 (%7,69)
BMI	28 (±4)	28 (±5)	31 (±5)	30 (±6)
Biopsy by D/C	52 (%72,22)	183 (%81,33)	58 (%85,29)	105 (%89,74)
Biopsy by vacuum	20 (%27,78)	42 (%18,67)	10 (%14,71)	12 (%10,26)
Hysterectomy after biopsy	26 (%36,11)	85 (%37,78)	29 (%42,65)	55 (%47,01)
WHO2020 Pre-malign/Malign	0 (%0)	9 (%4)	5 (%7,35)	27 (%23,07)

ET: Endometrial Thickness, HT: Hypertension, DM: Diabetes Mellitus, BMI: Body Mass Index, D/C: Dilatation and Curettage

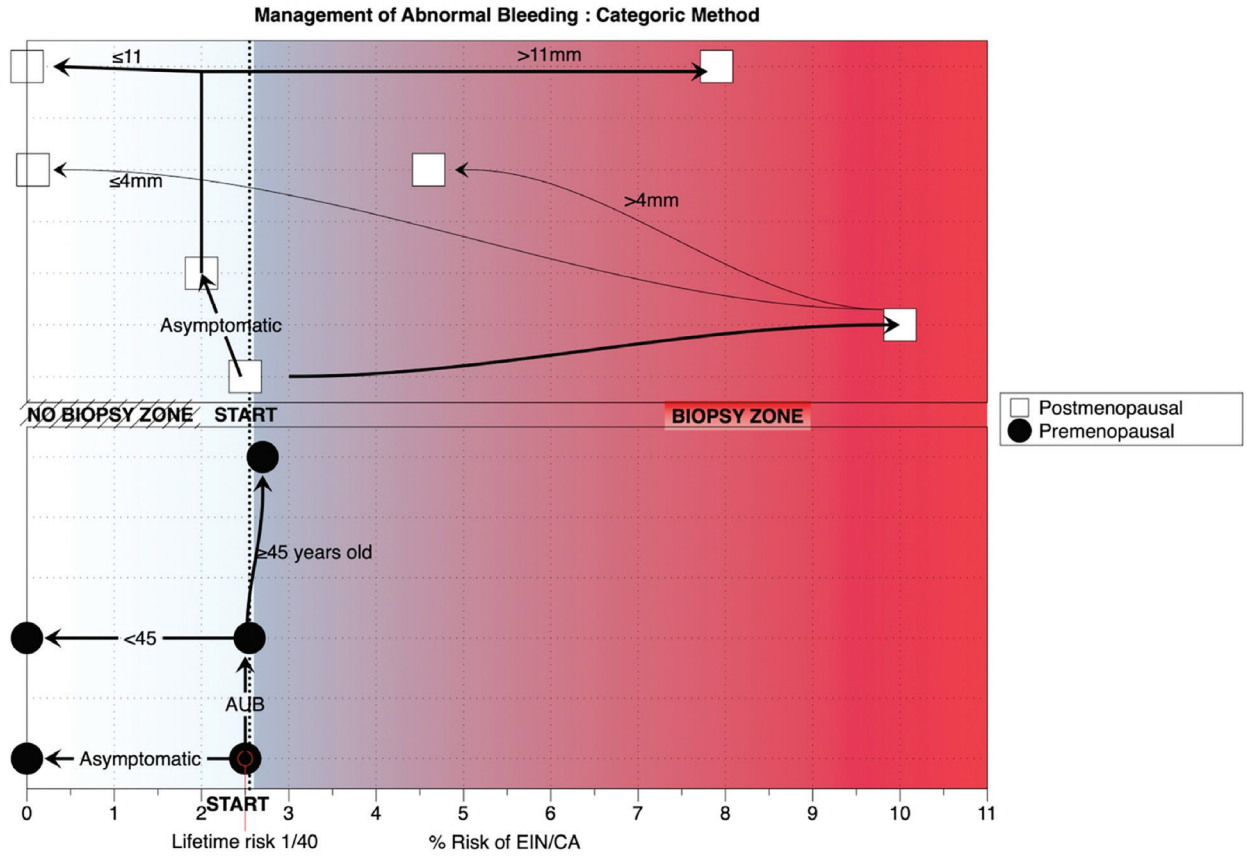


Figure 3: Odds ratio and confidence intervals of predictors in univariate analyses

Table 2. Results of Univariate and Multivariate Analyses.
Bootstrapping confidence intervals were similar, indicating internal validation of the study. ET: endometrial thickness

WHO2020 Classification	Symptomatic										Asymptomatic									
ET (mm)	35 Age	40 Age	45 Age	50 Age	55 Age	60 Age	65 Age	70 Age	75 Age	80 Age	35 Age	40 Age	45 Age	50 Age	55 Age	60 Age	65 Age	70 Age	75 Age	80 Age
2	0,4	0,6	1	1,5	2,4	3,7	5,6	8,4	12,4	18,2	0	0,1	0,1	0,2	0,2	0,4	0,6	0,9	1,4	2,2
3	0,5	0,8	1,2	1,9	2,9	4,4	6,6	9,9	14,6	21	0,1	0,1	0,1	0,2	0,3	0,5	0,7	1,1	1,7	2,6
4	0,6	0,9	1,4	2,2	3,4	5,2	7,9	11,7	17,1	24,3	0,1	0,1	0,1	0,2	0,4	0,6	0,9	1,3	2,1	3,2
5	0,7	1,1	1,7	2,7	4,1	6,2	9,3	13,8	19,9	27,9	0,1	0,1	0,2	0,3	0,4	0,7	1	1,6	2,5	3,8
6	0,9	1,4	2,1	3,2	4,9	7,4	11	16,2	23,1	31,8	0,1	0,1	0,2	0,3	0,5	0,8	1,2	1,9	3	4,5
7	1,1	1,6	2,5	3,8	5,8	8,8	13	18,9	26,5	35,9	0,1	0,2	0,3	0,4	0,6	1	1,5	2,3	3,5	5,4
8	1,3	2	3	4,6	6,9	10,4	15,3	21,9	30,3	40,3	0,1	0,2	0,3	0,5	0,8	1,2	1,8	2,8	4,2	6,4
9	1,5	2,3	3,6	5,5	8,3	12,3	17,8	25,2	34,4	44,9	0,2	0,2	0,4	0,6	0,9	1,4	2,2	3,3	5,1	7,7
10	1,8	2,8	4,3	6,5	9,8	14,4	20,7	28,9	38,7	49,5	0,2	0,3	0,5	0,7	1,1	1,7	2,6	4	6	9,1
11	2,2	3,4	5,1	7,8	11,5	16,9	24	32,9	43,2	54,2	0,2	0,4	0,5	0,8	1,3	2	3,1	4,7	7,2	10,7
12	2,6	4	6,1	9,2	13,6	19,6	27,5	37,1	47,8	58,8	0,3	0,4	0,7	1	1,6	2,4	3,7	5,7	8,5	12,7
13	3,2	4,8	7,3	10,9	15,9	22,7	31,4	41,6	52,5	63,2	0,3	0,5	0,8	1,2	1,9	2,9	4,4	6,7	10,1	14,9
14	3,8	5,7	8,6	12,8	18,6	26,2	35,5	46,1	57,1	67,4	0,4	0,6	1	1,5	2,3	3,5	5,3	8	11,9	17,4
15	4,5	6,8	10,2	15	21,6	29,9	39,9	50,8	61,6	71,4	0,5	0,7	1,1	1,8	2,7	4,2	6,3	9,5	14	20,2
16	5,4	8,1	12,1	17,6	24,4	34	44,5	55,4	65,9	75	0,6	0,9	1,4	2,1	3,3	5	7,5	11,2	16,4	23,4
17	6,4	9,6	14,2	20,5	28,5	38,3	49,1	60	70	78,3	0,7	1,1	1,7	2,5	3,9	5,9	8,9	13,2	19,2	26,9
18	7,6	11,4	16,6	23,7	32,5	42,8	53,7	64,4	73,7	81,3	0,8	1,3	2	3,1	4,7	7,1	10,6	15,5	22,2	30,7
19	9,1	13,4	19,4	27,7	36,7	47,4	58,3	68,5	77,2	84	1	1,5	2,4	3,7	5,6	8,4	12,5	18,1	25,6	34,8
20	10,7	15,7	22,4	31	41,1	52,1	62,8	72,4	80,3	86,4	1,2	1,9	2,9	4,4	6,6	9,9	14,7	21,1	29,3	39,2
Risk for>%2,5model (p<0,001) cut-off %95CI	12mm 1,1-6,3	10mm 1,3-5,8	7mm 1,3-4,8	5mm 1,4-5	3mm 1,4-5,7	2mm 1,7-7,8	2mm 2,4-12,5	2mm 3,3-19,8	2mm 4,4-30,5	2mm 5,8-44,2	24mm 0,5-12,6	22mm 0,5-10,6	20mm 0,7-10,5	17mm 0,7-8,4	15mm 0,9-8,2	13mm 1-8,4	10mm 0,8-7,8	8mm 0,8-8,8	5mm 0,6-9	3mm 0,6-10,8

table 2. All predictors were included in the multivariate analyses. In multivariate analyses, body mass index, hypertension, cigarette smoking and diabetes did not contribute to the model and were excluded. Independent variables had only the main effect, and there was no important interaction effect to include in the model. We have detected multicollinearity between age and menopausal status. Therefore, two models for regression analysis were analysed. Finally, we have selected model 1 consisting of symptom status, endometrial thickness and age for the best risk predicting method for pre-/endometrium cancer. Automated forward stepwise selection with entry testing based on the significance of the score statistic, and removal testing based on the probability of a likelihood-ratio statistic based on the maximum partial likelihood estimates have also yielded the same model that we have found. Regression analysis coefficients, and confidence intervals and odds ratio of the independent predictors in multivariate analyses are shown in figure 3 and table 2.

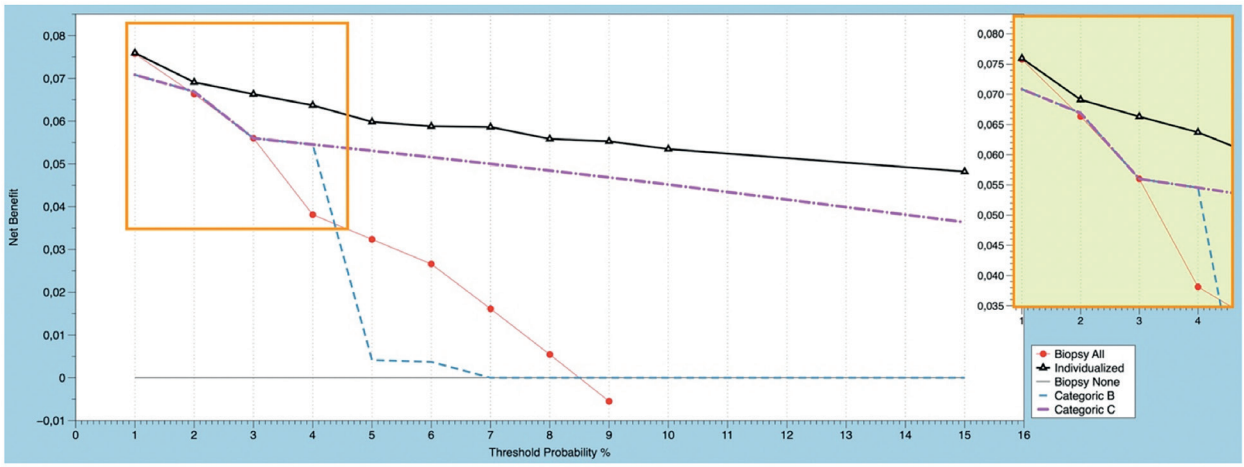
Diagnostic risk predictions and 95% confidence intervals were shown in table 2. Table 2 provides an insight into the predicted risk and confidence intervals for each endometrial thickness in mm and age for each 5-year group. This table provides a more user-friendly information for the daily practitioner rather than regression results. 3% risk was colored differently to create a visual summary.

Primary Outcome 1.2; Bootstraping

Bootstrapping had yielded similar confidence intervals in regression analyses (Table 2). A popular rule in clinical prediction modelling is the “one in ten rule” for sample size. According to this rule, one variable can be considered in a model for every 10 events (13). An event rate of 41 in our study can be adequate for a model with 3 predictors. The risk of pre/endometrial cancer in the study group was 8.5% and the mean estimated risk of the individualized model was 8.5±1.6. Brier score was 0.000002 indicating a near-perfect performance of the model.

Primary Outcome 1.3; Clinical Usefulness and Comparison to Current Practice

It is supposed theoretically that a model with better discrimination and calibration guides clinical judgement better but they are only statistical measures and can fall short to evaluate whether the risk model improves clinical decision making (14).



Default strategies were biopsy all patients and biopsy none. The graph gives the expected net benefit per patient relative to biopsy none.

Figure 4: Decision Curve Analyses for the individualized and comparator categoric models (Figure 1) to predict pre/invasive endometrial cancer

Decision curve analysis (Figure 4) showed the net benefit of the individualized model and comparator categoric models (model B and Model C). The net benefit of treat none is always 0 because this strategy has no true or false positives. Decision to biopsy everyone was also plotted at all reasonable thresholds. Treat all and treat none on the plot were default strategy. A model was regarded clinically useful at a specified threshold if it had a higher net benefit than treat all and treat none. If a model had a lower net benefit than any default strategy, we considered the model as harmful (15).

Categoric Model B and categoric model C (Figure 2 and Figure 4) have a higher net benefit than both treat all and treat none only for threshold probabilities between 3%-4.5% and above 3%, respectively.

The individualized model had a higher net benefit than default strategies above 1% threshold. The net benefit of models and default strategies are shown in detail for threshold probabilities 0-3.5% (Figure 4). When comparing two models, we have checked which model had the highest net benefit (14). The individualized model had a higher net benefit than model B and model C across the entire range of threshold probabilities.

Secondary Outcome 2.1; Discriminative Properties (AUC)

Discrimination assesses how well the model differentiates between those patients who experience the outcome and those who do not. AUCs of the individualized model and comparator AUROC model with symptom status, menopause and endometrial thickness > 4mm are shown in figure 5. Abnormal uterine bleeding/postmenopausal bleeding had an AUC of 59. In the IRPM endometrial thickness and age was used in conjuction with symptom staus and had an AUC of approximately 90. However, AUC of comporator models categoric predictors menopause and endometrial thickness <4mm was found to be 77.

Secondary Outcome 2.2; Change in ΔAUC,Violin Plots

The addition of age to symptom status and endometrial thickness in IRPM and addition of 4 mm cut-off to symptom status and menopausal status in the categoric model slightly increases the AUC but the magnitude of ΔAUC may fail to recognize some promising predictors (16). This is a common paradox where the base AUC is large (0.88 for symptom and endometrial measurement and 0.77 for symptom and menopause). (Figure 5). While the increase in AUC may be important, the small magnitude of the increase may lead to questioning its clinical significance (16). Therefore, we have delved into the density of probabilities in benign and pre/cancer group with violin plots (Figure 6).

In violin plots, when the convergence of shape in benign and pre/cancer groups are less, the model is deemed as more discriminative. The importance and contribution of the addition of age are evident in figure 6. When age is added to the model, it distributes the risks and provides a better discrimination. However, on the left side of the panel presenting the categoric model, discriminative property advanced slightly by the addition of 4 mm cut-off for endometrial thickness.

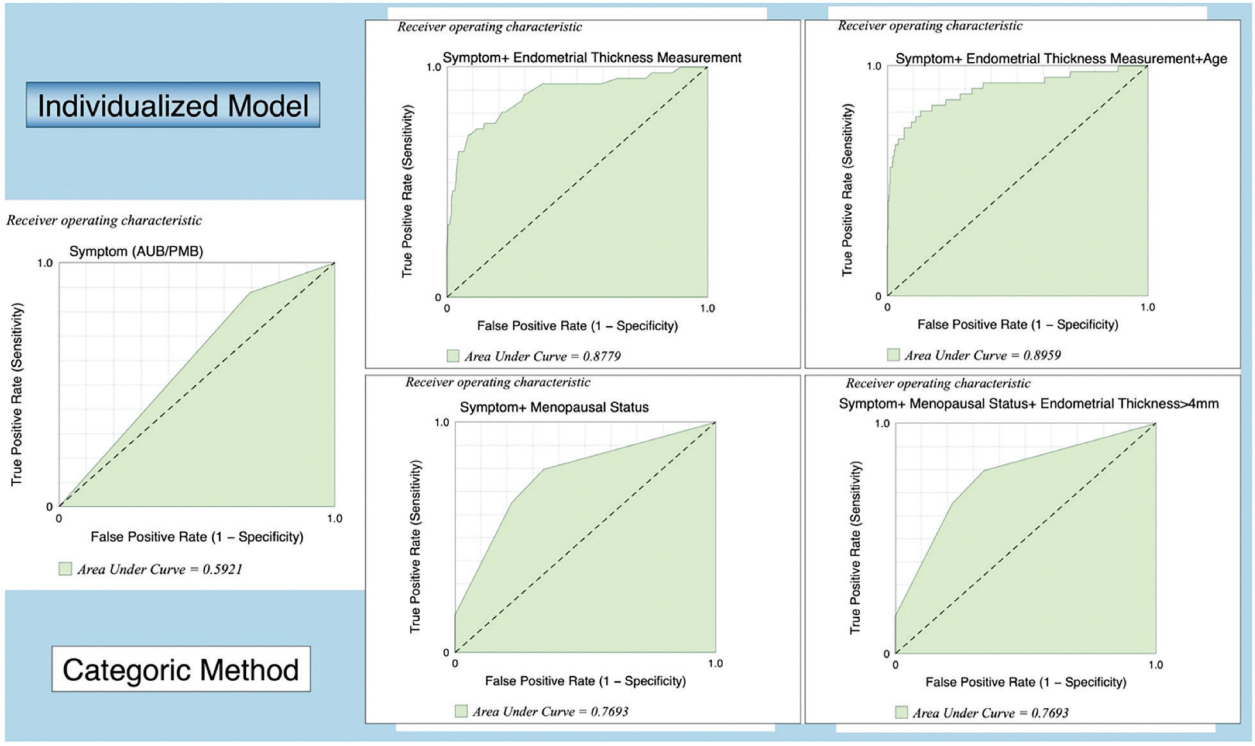


Figure 5: AUROC of individualized model and categoric model including abnormal bleeding/postmenopausal bleeding, menopause and endometrial thickness >4mm

Discriminative slopes, difference between median of models with and without a predictor gives also an insight (Figure 6); discriminative slopes using medians of the individualized model without and with age were 0.22 and 0.46, respectively. Discriminative slopes of the categoric model without and with were 0.20 and 0.32, respectively. The individualized model has a better discrimination slope.

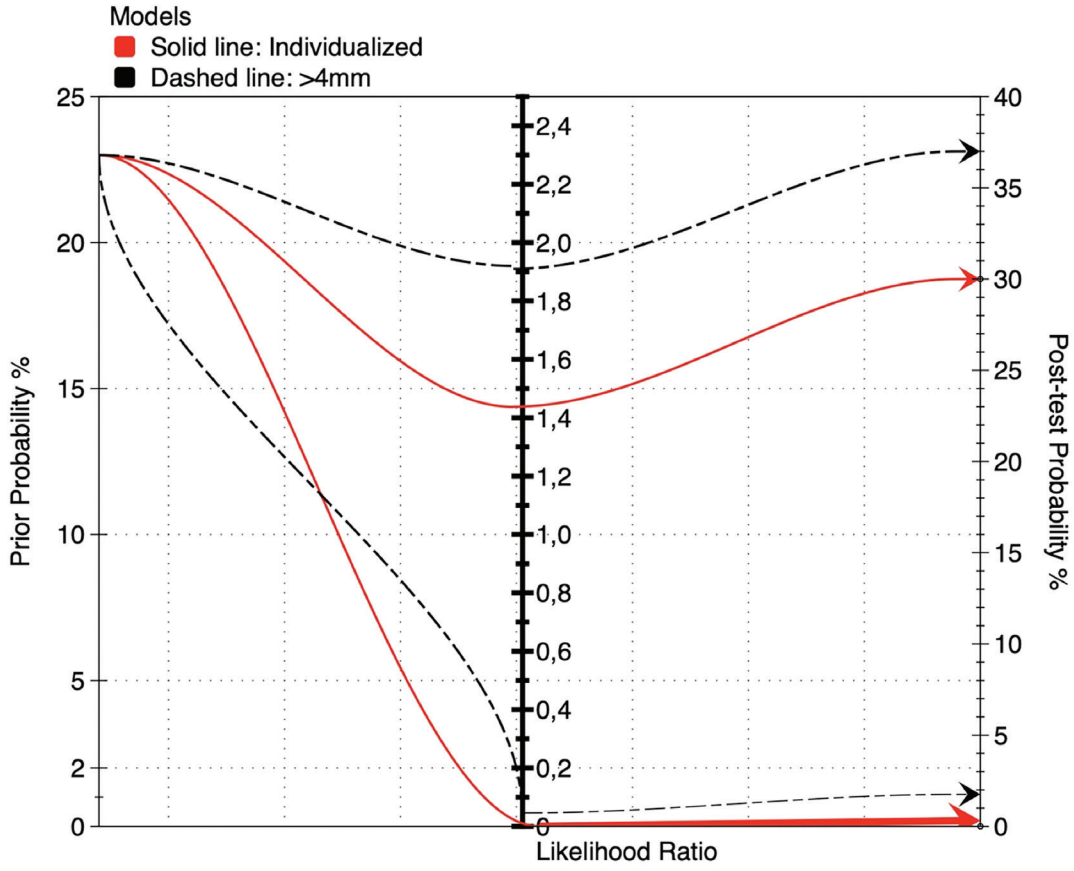
The effects of choosing different thresholds can also be viewed in figure 6. 3% risk threshold for biopsy is drawn as an example in the figure but one can increase or decrease the threshold. The area of violin plots of pre/cancer group under the threshold shows missed cases, whereas the area of violin plot of benign group shows unnecessary biopsies. Missed cases in the IRPM are less for every threshold, compared to the categoric method, and as the threshold is increased this becomes more evident because of the density of probabilities.

Secondary Outcome 2.3; Likelihood Ratios, Negative and Positive Predictive Values

For a threshold of 3% risk to trigger biopsy, sensitivity, specificity of the individualized model was 0.90, 0.63. Positive likelihood ratio was 2.44 (95% CI: 2.08-2.86). Negative likelihood ratio was 0.15 (95% CI: 0.06-0.39). Prior probability (odds) of 9% had changed to a posterior probability 18% (95% CI: 16%-21%) after a positive test (~ 1 in 5.4 with positive test had pre/cancer). Posterior probability after a negative test was 1% (95% CI: (1%-3%).

Discussion

Age is a significant risk factor that can be utilized for the prediction of endometrial cancer and EIN. The combination of age with endometrial thickness and symptom status provides a reliable prediction model and it is also a simple model. Simple models are easier to interpret and to use in practice (17). Current guidelines cover only a proportion of women; who are postmenopausal and have bleeding. Abnormal bleeding is a common complaint above women 35 years old however current guidelines does not cover these patients, even though they might have the risk of EIN/Endometrial cancer. Our model has a good discrimination and predicts the risk of pre/cancer in a heterogenous group. Addition of age to abnormal bleeding and endometrial thickness helps to identify patients in risk and increases the discriminative ability (Figure 6). The net benefit and clinical utility of the personalized risk model is better than the model which categorizes patients into groups (Figure



Individualized model with age and without age is shown on the left panel.Categoric model with and without endometrial thickness (ET) is shown on the right side. In the box plot, medians are written and discriminative slopes can be appreciated. As an example 3% threshold for biopsy is drawn with a solid line. The missed cases in the individualized model is less for every threshold, compared to the categoric method, and as the threshold is increased this becomes more evident because of the shape of distribution densities.

Figure 6: Distribution of predictions in benign (benign and hyperplasia without atypic) and pre/malign (endometrial intraepithelial neoplasia (EIN)/atypical hyperlasia and invasive endometrial cancer) groups.

4). The groups created in the categoric model are not uniform and involves patients having a wide spectrum of risk (Figure 6).

Importance of This Study and Mean Findings

We have found that age is an important predictor in univariate and multivariate analyses and bootstrap results. While the confidence interval of abnormal bleeding was too wide, using age and endometrial thickness provided a more precise and reliable prediction. The sample size of the study can be accepted as sufficient for conclusion. Net benefit and decision curve analysis showed that the individualized model is clinically useful above 1% threshold of pre-/endometrium cancer risk. The comparator models were built by the current best evidence and usually employed first as categorizing the patient based on symptom and menopausal status then stratifying by age, or endometrial thickness (Figure 2). These comparator models mimicing everyday practice and guidelines were inferior to the IRPM for every threshold (Figure 4). Our study also yielded clinically useful information about all models. The comparator models using the algorithms in Figure 2 were only useful above a threshold of 3% and below this threshold, it is found to be harmful to use the comparator model anyway.

Secondly, we have studied the discriminative properties of the individualized model and comparator model with AUC and violin plots. AUC of abnormal bleeding was 59. Figure 4 shows that the AUC of abnor-

mal bleeding with endometrial thickness is already high (AUC:88) in the IRPM. Addition of age to this model slightly increases the AUC (AUC:90). Figure 4 shows that the AUC of abnormal bleeding and menopause in the comparator model was already high (AUC:77). Addition of endometrial thickness >4mm cut-off only increased it to 78. However, it would be incorrect to conclude that age and >4 mm cut-off have minimal effect on the discriminative ability of each model. The main effect that can be desired in these models with a large AUC is addition of a factor that helps to discriminate better and decrease the rates of missed cases and unnecessary biopsies. Violin plots yielded the density of risk prediction for each model and concluded the benefit of addition of age or endometrial thickness >4mm to the models. Particularly in the IRPM, the addition of age helped to discriminate the patients with risk for pre/cancer.

Findings and Comparison in Literature

There is a paucity in the literature regarding risk prediction models and their usage to triage women for further biopsy. We could not identify any study including women above 35 years old and assessing the clinical usefulness as net benefit without splitting the database into premenopausal and postmenopausal. Alblas et al could identify only two researches for risk prediction out of 2756 references (18). Most of the other articles were performed to analyse discriminative diagnostic approaches. They found only 8 studies that described the development of models for symptomatic women (diagnostic models). The findings of these and other studies were comparable with our secondary outcomes. Seven of these studies included only women with postmenopausal bleeding (18) and only three studies included women aged 40 years with ultrasound endometrial abnormalities (19-22). These three studies were comparable with our tertiary outcomes.

The epidemiological risk prediction models were built to identify risk factors for endometrial cancer and predict the risk of endometrial cancer in 5 years or 20 years (23, 24) these studies aimed to identify high-risk women to take preventive measures such as physical activity or weight loss. Hüsler et al reported that a large part of the overall discrimination capacity is based on a woman's age and body mass index, smoking, reproductive history and hormone replacement therapy were other substantial risk factors. We have also found that age was an important risk predictor and it could be used to triage patients because it helps to individualize the predicted risk. Similar to other studies, body mass index, hypertension and non-smoking were significant diagnostic risk predictors in univariate analyses of our study.

Studies Evaluating the Discriminative Diagnostic Models

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 and various ultrasonographic measurements and biomarkers. Madkour et al enrolled 60 women with postmenopausal bleeding who had an endometrial thickness >5mm and analyses a model using International Endometrial Tumor Analysis group ultrasonography criteria (20). This risk model had an accuracy of about 95% and an area under the curve 0.9. Opolskiene et al reported that a model including age, use of warfarin and endometrial thickness had an AUC of 0.82 (95% CI, 0.76–0.87) for women with postmenopausal bleeding (21). They have found that the addition of endometrial thickness and doppler indices may increase the discriminative property. Gianella et al studied a risk-Scoring Model for the prediction of Endometrial Cancer among Symptomatic Postmenopausal Women with Endometrial Thickness > 4 mm (22). The best predictors of endometrial cancer were recurrent vaginal bleeding (odds ratio), the presence of hypertension endometrial thickness > 8 mm, and age > 65 years. Angelo et al assessed a risk stratification tool of endometrial cancer, combining serum markers, clinical and ultrasound characteristics. They found that preoperative age, symptom, HE4 levels, and ultrasound endometrial thickness were found statistically significant, and were included in a multivariate logistic regression model to determine the probability to have endometrial cancer. They reported AUC 0,91-0.95, 89-93.3% sensitivity and 95-97.1% specificity positive predictive value 0,73-0,83; negative predictive value 0.98.

The major drawback of the overall studies was in the methodology. Among these studies enrolling only symptomatic postmenopausal women, all except one assumed that all women with endometrial thickness lower than the accepted cut-off 4mm did not have pre/cancer. Besides, the cut-off values for endometrial thickness are controversial. Patel et al studied the histopathological results of women who are older than 55 years old and transvaginal endometrial thickness measurement (25). They have concluded that cut-off should be lower than anticipated and cut-off value of 3mm would provide 96.9% sensitivity and the most reliable predictor was found to be postmenopausal bleeding. Clarke et al reported that women with postmenopausal bleeding have the same risk of pre/invasive endometrial cancer (12) and they concluded that 4mm cut-off should be used only in women younger than 60 years old. These

findings were also consistent with our results. The cut-off value of the endometrial thickness should be lower as age progresses as suggested by our findings (Table 2). Gupta et al reported in their meta-analyses that a positive test result raised the probability of carcinoma from 14.0% (95% CI 13.3–14.7) to 31.3% (95% CI 26.1–36.3), while a negative test reduced it to 2.5% (95% CI 0.9–6.4). They had concluded ultrasound measurement of endometrial thickness alone, using the best–quality studies cannot be used to accurately rule even in women with postmenopausal bleeding (26). Recent analyses revealed that sensitivity, positive predictive value and specificity of endometrial thickness with a cut off value ≤4mm has been reported to be lower than the previous reports (27). Using a cut-off value in asymptomatic postmenopausal and premenopausal women is more debatable.

There is no consensus on the use of endometrial thickness in asymptomatic postmenopausal women. The study of Smith-Bindman et al suggested using a threshold value of the endometrial thickness of ≥11m for biopsy in postmenopausal asymptomatic women (13). However, Breijer et al found in their meta-analyses that no threshold value should be used in these women. A risk prediction model developed only for postmenopausal women with bleeding would not be ideal for everyday practice and such a design would miss premenopausal and postmenopausal asymptomatic patients. These missed cases could impair the diagnostic accuracy of these models and put their discriminative ability into question. Calibration and discrimination are important aspects of a prediction model, however, these aspects do not assess the clinical usefulness (28).

Advanced ultrasonography techniques are discussed in the literature however they can be time-consuming, costly and can be universally available only in limited settings.We aimed to analyze a model in a general dataset of women including premenopausal as well as postmenopausal women. The model developed should be simple and preferably should not add to the cost. A complex model involves many confounders and can not be practical to use. Our final multivariate model by abnormal bleeding status, endometrial thickness and age are practical to use with minimal multivariate predictors. It doesn't add to the cost such as a biomarker. Above all, the net benefit of the model was higher compared to other models (Figure 4) and allows for individualization of the biopsy decision.

Strength and Limitations in Comparison With Literature

The probability of non-replication of published studies with p-values in the range 0.005 to 0.05 is roughly 0.33 and non-reproducibility in non-randomized studies can be as high as 80% to 90% (29,30). The probability of rejecting the null hypothesis is at least 23% and typically closes to 50% at a p-value of 0.05. In previous studies, predictors were selected based on invariable analyses with a p-value of 0.20 or p-value of 0.05 and statistical significance level of 0.05 were used however we have used more stringent p values for statistical analyses. A p-value of 0.005 was chosen in our study because it allows for good reproducibility of the study.

Another strength of our study was that interobserver and intraobserver variability was eliminated and all patients had reference test regardless of the index test. In the majority of previous studies, a biopsy(index test) was not performed below a certain cut-off value and the models were analyzed only in a subset of women such as postmenopausal. Besides, ultrasound, D&C and target was evaulated by different performers with various experience. The difference of endometrial thickness suggested by various guidelines is within 1mm which is subject to intra- and inter-observer variability in the index test. Karlson et al compared the endometrial thickness measurement between experienced and inexperienced sonographers and found that there was about 1.5mm of mean discrepancy from true measurement for the inexperienced sonographers. Dueholm et al reported that findings of transvaginal ultrasonography, gel infusion sonography and hysteroscopy were not reproducible for postmenopausal patients with abnormal bleeding when the endometrial thickness was above 4 mm. It can be argued that our sample size can be larger. To analyze all the predictors in univariate analyses, an event of 100-200 and at a prevalence of 10% at least a sample size of 1200 could be enough. Our sample size is enough for 3-4 predictors, that were included in the multivariate model. We have performed bootstrapping and internally validated the study. A major limitation of our study is that it was not externally validated.

Another point of argument can be the biopsy technique used. It may be argued that hysteroscopy can be used instead of dilatation curettage. There is one randomized study that compares hysteroscopy with blind biopsy in women with postmenopausal bleeding and an endometrial thickness greater than 4mm. In this study, hysteroscopy was found to be superior to blind biopsy to detect EIN/endometrial cancer at the endometrial polyps. However, this randomized study was underpowered (31). Besides hysteroscopy was compared with pipelle biopsy rather than D&C. Pipelle biopsy has a high false-negative rate. To the authors' knowledge, there is no randomized prospective study comparing the dilatation and curettage with hysteroscopy and it is not

possible to conclude one is superior to the other. In addition, there are also some concerns about hysteroscopy. Hysteroscopy can also cause to disseminate endometrial tumor cells to the peritoneal cavity (32). Although the significance of this dissemination is unknown for endometroid endometrial cancer, it can be important for other histologic subtypes such as carcinosarcoma. Aside from the fear of dissemination, the lesion severity and where to take targeted biopsy by office hysteroscopy is not well established. Hysteroscopy seems useful to diagnose polyps and other endometrial pathologies rather than pre-/cancer. Currently, it is not possible to identify atypical endometrial hyperplasia or concurrent cancer tissue by patterns seen in hysteroscopy but benign pathologies such as polyps, submucous myomas can be differentiated. Hysteroscopy and simultaneous D&C are the gold standard when the procedure is planned under anesthesia. We have performed dilatation and curettage under general anesthesia to all women and in case of severe cervical stenosis that prevents cervical dilatation >8 Hegar dilators, we performed high vacuum curettage. Furthermore, one of the limitations of our study that using of WHO 94 and EIN classification together in the pathological evaluation phase.

Interpretation of Results and Clinical Implications

Combining age with ultrasonographic and clinical information allows individualization of the decision process. The model built in this study is a well-calibrated risk model with moderate to good discriminatory accuracy and can aid in individual decision-making. The IRPM is superior to biopsy all patients and the comparator model using abnormal bleeding, menopause and a cut off of endometrial thickness >4mm. The comparator model mimicking current guidelines and practice based on literature is only useful above a threshold of 3% and was found to be harmful under this threshold. Using the model found in this study, the clinician can discuss the individualized risks with the patient and the patient can also contribute to the decision process better informed. In patients with important co-morbidities, this risk-based system also allows making a comparison if the risk of cancer outweighs the risk of co-morbidities concerning the procedure.

Endometrial cancer prevalence and risk factors change geographically (33) therefore categorizing patients into general groups and using cut-off values may render the cut-off values ineffective. Our model can be adjusted or recalibrated for local circumstances by combining information captured in the original model with information from new individuals and the updated model can have improved transportability to other individuals in new settings (34).

Conclusion

Women with endometrial cancer are from a heterogeneous population involving premenopausal and postmenopausal women. The prediction model combining abnormal bleeding, age and endometrial thickness may increase the sensitivity and specificity of detecting pre/endometrial cancer in this heterogeneous group. We were able to compare the individualized risk models' superiority to other models using abnormal bleeding, menopause and endometrial thickness >4mm. The value of diagnostic discriminatory cut-offs, suggested by guidelines is lower than expected in a heterogenous group. Although the model developed in this study can be updated and calibrated for different populations, this approach is cumbersome for cut-off values. IRPM does not need any additional costs and time-consuming analysis and may aid the patient to contribute to the decision of further investigation.

Data Sharing

The datasets generated and analysed during the current study are not publicly available because of the informed consent documents. The data of the findings of this study can be obtained from the last author on reasonable request and with the permission of the contributing cancer registries of the university.

Roles of Authors:

OKK had contributed to the collection of data, study design and medical writing as first investigator. EE was the methodological chief and contributed to study desgin, medical writing and statistical analyses. EE and KKB was the principal investigators that had undertaken the clinical tests. VO and IT had contributed to collection of data and follow up of patients and medical writing.

References

1. Henley SJ, Miller JW, Dowling NF, Benard VB, Richardson LC. Uterin Cancer Incidence and mortality-United states, 1999-2016. Morbidity and Mortality Weekly Report 2018; 67: 1333-38.
2. Pennant ME, Mehta R, Moody P, et al. Premenopausal abnormal uterine bleeding and risk of endometrial cancer. BJOG 2017; 124: 404-11.
3. Emery J, Vedsted J. New NICE guidance on diagnosing cancer in general practice. Br J Gen Pract 2015; 65: 446-47.

4. Ioannidis JP, Haidich AB, Pappa M, et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. JAMA 2001; 286: 821-30.
5. Wong ASW, Lao TTH, Cheung CW, et al. Reappriasal of endometrial thickness for the detection of endometrial cancer in postmenapausal bleeding, a retrospective cohort study. BJOG 2016; 123: 439-46.
6. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ 2015; 351: h5527.
7. Initiative S. STROBE checklist for cohort, case- control, and cross-sectional studies. 2007. https://www.strobe-statement.org/download/strobe-checklist-wide-cohort-case-control-and-cross-sectional-studies-combined-pdf (accessed 17 Dec, 2021).
8. Ellenson LH, Matias-Guiu X, Mutter GL. Endometrial hyperplasia without atypia. In Female Genital Tumours WHO Classification of Tumours Series, 5th Edition, Volume 4. IARC Press, Lyon, 2020; 248-49.
9. Lax SF, Mutter GL. Endometrial atypical hyperplasia/endometrioid intraepithelial neoplasia. In Female Genital Tumours WHO Classification of Tumours Series, 5th Edition, Volume 4. IARC Press, Lyon, 2020; 250-51.
10. Sobczuk K, Sobczuk A. New classification system of endometrial hyperplasia WHO 2014 and its clinical implications. Prz Menopauzalny 2017; 16: 107-11
11. Bindman RS, Weiss E, Feldstein V. How thick is too thick? When endometrial thickness should prompt biopsy in postmenopausal women without vaginal bleeding. Ultrasound Obstet Gynecol 2004; 24: 558-65.
12. Clarke MA, Long BJ, Sherman ME, et al. Risk Assessment of endometrial cancer and endometrial intraepithelial neoplasia in women with abnormal bleeding and implications for clinical management algorithms. Am J Obstet Gynecol 2020; 223: 549.
13. Huang Y, Lİ W, Macheret F, Gabriel RA, Machado LO. A tutorial on calibration measurements and calibration models for clinical prediction models. Journal of the American Medical Informatics Association 2020; 27: 621-33.
14. Calster BV, Vickers AJ. Calibration of risk prediction models: impact on decision-analytic performance. Med Decis Making 2015; 35: 162–9.
15. Calster BV, Wynants L, Verbeek JFM, et al. Reporting and Interpreting Decision Curve Analysis: A Guide for Investigators. Eur Urol 2018; 74: 796-04.
16. Pencina MJ, D'Agostino RB, Vasan RS. Statistical methods for assessment of added usefulness of new biomarkers. Clinical Chemistry and Laboratory Medicine 2010; 48: 1703-11.
17. Chawdhury MZI, Turin TC. Variable selection strategies and its importance in clinical prediction modelling. Fam Med Com Health 2020; 8: e000262.
18. Alblas M, Velt KB, Pashayan N, Widschwentter M, Steyerberg EW, Vergouwe Y. Prediction models for endometrial cancer for the general population or symptomatic women: A systematic review. Critical Reviews in Oncology/Hematology 2018; 126: 92-99.
19. 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: 5733-39.
20. Madkour NM. An ultrasound risk-scoring model for prediction of endometrial cancer in post-menopausal women (using IETA terminology). Middle East Fertility Society Journal 2017; 22: 201-05.
21. Opolskiene G, Sladkevicius P, Valentin L. Prediction of endometrial malignancy in women with postmenopausal bleeding and sonographic endometrial thickness ≥ 4.5 mm. Ultrasound in Obstetrics and Gynecology 2011; 37: 232-40.
22. Giannella L , Mfuta K, Setti T, Cerami LB, Bergamini E, Boselli F. A Risk-Scoring Model for the Prediction of Endometrial Cancer among Symptomatic Postmenopausal Women with Endometrial Thickness > 4 mm. Biomed Res Int 2014; 2014: 130569.
23. Husing A, Dossus L, Ferrari P, et al. An epidemiological model for prediction of endometrial cancer risk in Europe. Eur J Epidemiol 2016; 31: 51-60.
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: e1001492.
25. Patel V, Wilkinson EJ, Chamala S, Lu X, Castagno J, Rush D. Endometrial Thickness as Measured by Transvaginal Ultrasound and the Corresponding Histopathologic Diagnosis in Women With Postmenopausal Bleeding. Int J Gynecol Pathol 2017; 36: 348-55.
26. Gupta JK, Chien PFW, Voit D, Clark TJ, Khan KS. Ultrasonographic endometrial thickness for diagnosing endometrial pathology in women with postmenopausal bleeding: a meta-analysis. Acta Obstet Gynecol Scand 2002; 81: 799-16.
27. Saccardi C, Vitagliano A, Marchetti M, et al. Endometrial Cancer Risk Prediction According to Indication of Diagnostic Hysterescopy in Post – menopausal Women. Diagnostics(Basel) 2020; 10: 257.

28. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur Heart J 2014; 35: 1925-31.
29. Boos DD, Stefanski LA. P-Value Precision and Reproducibility. Am Stat 2011; 65: 213–21.
30. Young S. Everything is Dangerous: a Controversy. National Institute of Statistical Sciences. 2008. https://www.openphilanthropy.org/sites/default/files/Stanley%20Young%20slides%20on%20multiple%20testing%201.pdf (accessed 7 Dec, 2021).
31. Hanegem N, Breijer MC, Slockers SA, et al. Diagnostic workup for postmenopausal bleeding: a randomised controlled trial. BJOG 2017; 124: 231–40.
32. Obermair A, Geramou M, Gucer F, et al. Does hysteroscopy facilitate tumor cell dissemination? Incidence of peritoneal cytology from patients with early stage endometrial carcinoma following dilatation and curettage (D & C) versus hysteroscopy and D & C. Cancer 2000; 88: 139-43.
33. Clarke MA, Long BJ, Morillo ADM, Arbyn M, Bakkum-Gamez JN, Wentzensen N. Association of Endometrial Cancer Risk With Postmepausal Bleeding in Women. A Systematic Rewiev and Meta-analiysis. JAMA intern Med 2018; 178: 1210-22.
34. Shipe ME, Deppen SA, Farjah F, Grogan EL. Developing prediction models for clinical use using logistic regression: an overview. J Thorac Dis 2019; 11: 574–84.