Management of endometrial abnormalities in postmenopausal women, an individualized approach
Breijer, M.C.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
External validation of a mathematical model to estimate the probability of endometrial cancer in women with postmenopausal bleeding

M.C. Breijer
N. van Hanegem
B.C. Opmeer
H.C. van Doorn
L. Valentin
T. van den Bosch
R.H.M. Verheijen
B.W.J. Mol
A. Timmermans

Submitted
ABSTRACT

**Objective**: To externally validate the previously developed ‘patient characteristics’ and ‘patient characteristics and transvaginal sonography (TVS)’ models estimating the probability of endometrial cancer in women with postmenopausal bleeding (PMB).

**Design**: External validation study.

**Setting**: Two independent datasets of consecutive women not using hormone replacement therapy with a first episode of PMB.

**Population**: One dataset was prospectively collected in three general hospitals in the Netherlands including 559 women, whereas the other dataset was prospectively collected in a university hospital in Sweden including 433 women.

**Methods**: We applied the two models on the two validation databases. The performance of the models was assessed by examining discrimination and calibration.

**Main Outcome Measures**: Discrimination is presented with a receiver operator characteristic (ROC) curve and an area under the ROC curve (AUC). Calibration is presented graphically with calibration plots.

**Results**: The AUC for the ‘patients characteristics’-model was 0.71 (95% CI 0.65 to 0.76) in the Dutch database and 0.69 (95% CI 0.64 to 0.73) in the Swedish database. The AUC for the ‘patient characteristics and TVS’-model was 0.89 (95% CI 0.86 to 0.92) and 0.89 (95% CI 0.86 to 0.91) in the Dutch and Swedish database, respectively.

**Conclusions**: Although both ‘patient characteristics and TVS’ models maintained their diagnostic performance in two independent validation databases, they did not seem to offer advantages over endometrial thickness measurement alone. The ‘patient characteristics’ model is able to select women with a low risk of endometrial cancer to be reassured without further testing, thus allowing its use in a setting without TVS.
INTRODUCTION

Postmenopausal bleeding (PMB) is a common complaint in postmenopausal women who present in both primary and secondary care. Immediately after menopause, PMB occurs in 10% of women. The main objective of the diagnostic work-up for women presenting with PMB is to rule out endometrial cancer. In women with PMB, the probability of endometrial cancer rises from 1% in women younger than 50 years to 24% in women older than 80 years and, regardless of age; the risk of malignancy is higher in women with obesity (18%) or diabetes (21%) than in women without these risk factors (8.0%).

In the 1990s endometrial thickness measurement with transvaginal ultrasonography (TVS) was introduced as a test to rule out endometrial cancer. However, there is debate on which cut-off value for endometrial thickness should be used when deciding if endometrial sampling needs to be performed or not. A post-test probability of endometrial cancer of 1% seems the worldwide-accepted threshold for patient reassurance. The post-test probability depends not only on test characteristics but also on the pre-test probability, which depends on patient’s characteristics. In women with PMB, characteristics that define the pre-test probability of endometrial cancer are: age, time since menopause, body mass index (BMI), hypertension, diabetes mellitus, and parity. Current guidelines are mainly based on sonographic endometrial thickness only and do not systematically take these additional characteristics into account.

In 2007 we reported on the development of two multivariable prediction models to estimate the probability of endometrial cancer in women with PMB. Details on the development of the two prediction models can be found in a previous publication. In short, data on 540 women with PMB, not using hormone replacement therapy (HRT), were included in a multivariable regression model. Patient characteristics that satisfied the criteria for inclusion in the model were age, BMI, diabetes, parity and the use of anticoagulants. Two models were developed: (1) the ‘patient characteristics’-model, including the five characteristics listed above and (2) the ‘patient characteristics and TVS’-model, including the five patient characteristics and endometrial thickness as measured by TVS. The area under the receiver operating characteristic curve (AUC) for the ‘patient characteristics’ model in the development database was 0.76 (95% CI 0.71 – 0.82) and for the ‘patient characteristics and TVS’ model 0.90 (95% CI 0.87 – 0.93).

For a successful implementation, a prediction model should be validated externally in an independent population. The aim of the present study was to externally validate the diagnostic performance of the existing two models in two independent prospectively collected datasets of women with PMB and to compare these models to the accuracy of endometrial thickness measurement alone.
METHODS

The development of the multivariable models

We developed two multivariable logistic regression models, which we described in the article by Opmeer et al.\textsuperscript{18} The first model only contains characteristics from the history of the women and is referred to as the ‘patient characteristics model’. The second model, an extension of the first model also includes endometrial thickness as measured with TVS and is referred to as the ‘patient characteristics and TVS model’. Categorical variables with subdivisions (e.g. type and management of diabetes) were dichotomised (e.g. diabetes: yes/no). Since we reported previously that the accuracy of endometrial thickness measurement was different in obese and non-obese women and in women with diabetes and those without,\textsuperscript{5} differences in diagnostic performance across sub-groups were evaluated through interaction terms. Further statistical details are provided in the original article.\textsuperscript{18}

Study population used for external validation

For the present study, we used two prospectively collected databases:

I. Dutch database: This database includes all women presenting with postmenopausal bleeding at the TweeSteden hospital Tilburg, the Maxima Medical Centre Veldhoven and the St. Antonius hospital Nieuwegein in the Netherlands between January 2009 and April 2011. Menopause was defined as at least one year of amenorrhea. No age criterion was used, if there were doubts about the postmenopausal status this was confirmed by hormone level testing. The following patient characteristics were recorded: age, years since menopause, BMI, parity, HRT use, hypertension, diabetes, use of anticoagulants and endometrial thickness as measured by TVS. If double endometrial thickness was $>4.0$ mm endometrial biopsy using the Pipelle\textsuperscript{®} (Laboratoire CCD, Paris, France) was performed. In case of a failed endometrial biopsy, hysteroscopy with directed biopsy was performed.

II. Swedish database: This database includes all women presenting with PMB at the Skåne University Hospital Malmö postmenopausal bleeding clinic between November 2002 and June 2009. Menopause was defined as at least one year of amenorrhea after the age of 40. The following patient characteristics were recorded: age, age at menopause, weight, height, parity, HRT use, hypertension, diabetes, use of anticoagulants and endometrial thickness as measured by TVS. If double endometrial thickness was $\geq 4.5$ mm, saline infused sonography (SIS) was performed. If there were no focal lesions in the uterine cavity at SIS, an endometrial sample using the Endorette\textsuperscript{®} (Medscan AB, Malmö, Sweden) was taken. If there were focal lesions at SIS or if SIS failed, hysteroscopic resection was performed.\textsuperscript{20}
All women in both databases were instructed to contact the hospital if recurrence of bleeding occurred. Follow-up in the Dutch database was based on data collected from case notes. For the purpose of this study, all women with an endometrial thickness below the cut-off value (without endometrial biopsy and without recurrent bleeding) were considered negative for endometrial cancer. In case of recurrent bleeding hysteroscopy was performed. In one patient endometrial cancer was diagnosed after recurrent postmenopausal bleeding during follow-up. The median follow-up time in this database was 26 months (range 18 to 43 months).

In the Swedish database, all women with an endometrial thickness $\leq 4.4$ mm (and therefore without a histological diagnosis of the endometrium) were matched with the regional cancer register to ascertain if any of these women were diagnosed with endometrial cancer after inclusion in the study.

Women were excluded from further analysis if the endometrial thickness was not measurable.

In line with the classification in the development database, definite disease state was determined as benign in women whose histology specimen showed atrophy, benign polyps, simple hyperplasia or proliferative endometrium. Premalignancy, defined as any form of hyperplasia with atypia, and malignancy in the histology specimen were combined in the analysis in the diagnostic group ‘endometrial malignancy’ since both diagnostic categories warrant further treatment.

**Statistical Analysis**

**Missing values**

Generally, dropping cases with missing values (complete case analysis) yields biased results, and the discriminative ability of a multivariable model is reduced when a case with missing values is dropped from the analysis.\textsuperscript{21} In multiple imputation, each missing value is imputed several times. The variation among the imputations reflects the uncertainty with which the missing values can be predicted from the observed ones. After combining the results, the overall estimates and standard errors reflect missing data uncertainty.\textsuperscript{22-24} In our validation study, we performed multiple imputations for missing variables, with separate imputation rounds for each of the two databases.

**External validation**

We applied the two models to the women in the Dutch and Swedish databases. We assessed the performance of the models by examining the calibration (agreement between predicted risks and observed frequencies of endometrial cancer) and discrimination (the ability of the models to distinguish between women with and without endometrial cancer). To assess calibration, we plotted the predicted probabilities of endometrial cancer and the observed
proportion of endometrial cancer by deciles of the predicted probabilities in a calibration plot.\textsuperscript{25} Calibration is considered perfect if the intercept is 0 and the calibration slope is 1.\textsuperscript{26,27} We assessed discrimination by calculating the area under the receiver operator characteristic curve (AUC). To compare the performance of the two models with the performance of the currently applied strategy of measuring endometrial thickness (‘TVS-only’), we calculated the AUC for ‘TVS-only’.

**RESULTS**

The two databases available for external validation consisted of 559 Dutch and 433 Swedish women with PMB not using HRT. Table 1 shows the characteristics of women in the two databases and the percentage of missing data per database. Age, time since menopause, anticoagulants use, body mass index (BMI), endometrial thickness and the frequency of endometrial cancer differed significantly between the two validation populations, women in the Swedish database being older, having lower BMI, thicker endometrium, a higher percentage of endometrial cancer and more women used anticoagulant therapy.

**Table 1.** Patient characteristics and missing values in validation databases.

<table>
<thead>
<tr>
<th></th>
<th>Swedish database</th>
<th>Missing</th>
<th>Dutch database</th>
<th>Missing</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>433</td>
<td>0</td>
<td>559</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>67.4 +/- 11.8</td>
<td>0</td>
<td>61.8 +/- 9.95</td>
<td>0</td>
<td>&lt;0.01#</td>
</tr>
<tr>
<td>DM</td>
<td>66 (15.2)</td>
<td>0</td>
<td>72 (12.9)</td>
<td>1 (0.2)</td>
<td>0.29†</td>
</tr>
<tr>
<td>HT</td>
<td>176 (40.6)</td>
<td>0</td>
<td>196 (35.1)</td>
<td>0</td>
<td>0.07†</td>
</tr>
<tr>
<td>AC</td>
<td>88 (31.9)</td>
<td>157 (13.2)</td>
<td>94 (16.8)</td>
<td>0</td>
<td>&lt;0.01†</td>
</tr>
<tr>
<td>BMI (kg/m(^2))</td>
<td>27.8 +/- 6.4</td>
<td>24 (5.5)</td>
<td>30.2 +/- 8.2</td>
<td>260 (43.4)</td>
<td>&lt;0.01#</td>
</tr>
<tr>
<td>TMP (years)*</td>
<td>16 (5-26)</td>
<td>6 (1.4)</td>
<td>5 (2-14)</td>
<td>124 (22.2)</td>
<td>&lt;0.01‡</td>
</tr>
<tr>
<td>Nulliparity</td>
<td>45 (10.6)</td>
<td>8 (1.8)</td>
<td>60 (13.5)</td>
<td>114 (20.4)</td>
<td>0.19†</td>
</tr>
<tr>
<td>ET (mm)*</td>
<td>6.0 (3.2-13.0)</td>
<td>NA</td>
<td>5.7 (2.5-10.0)</td>
<td>NA</td>
<td>0.02‡</td>
</tr>
<tr>
<td>Endometrial (pre-) malignancy</td>
<td>65 (15.0)</td>
<td>3 (0.69)</td>
<td>57 (10.2)</td>
<td>7 (1.3)</td>
<td>0.02†</td>
</tr>
<tr>
<td>Atypical hyperplasia</td>
<td>62 (14.3)</td>
<td>0</td>
<td>50 (8.9)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Endometrial cancer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

DM, diabetes mellitus; HT, hypertension; AC, anticoagulants; BMI, body mass index; TMP, time since menopause; ET, endometrial thickness; NA, not applicable.

Results presented are n (%), mean +/-SD or median and (interquartile range)

* not normally distributed, values presented as median and (interquartile range)

# Independent samples T-test

† Chi-square

‡ Independent samples Mann-Whitney U Test
Calibration plots for the ‘patient characteristics’ model and the ‘patient characteristics and TVS’ model in the Dutch and Swedish database are presented in Figure 1. In the Dutch database, the calibration slope is better for the ‘patient characteristics and TVS’ model compared to the ‘patient characteristics’ model. In patients with a high risk of endometrial carcinoma predicted probabilities are slightly overestimated when using the ‘patient characteristics model’. Because this is only in the high-risk region, the model is capable of selecting women with a low risk, who can be reassured without further invasive testing.

Figure 1. Calibration plots ‘Patient characteristics only’-model and ‘Patient characteristics and TVS’-model.
In the Swedish database, the calibration slope performed better for the ‘patient characteristics’ model compared to the ‘patient characteristics and TVS’ model. For the ‘patient characteristics’ model, predicted probabilities were close to the observed frequency in all patients, low- or high risk. The ‘patient characteristics and TVS’ model underestimated the probability of endometrial cancer over almost the whole range of probabilities except for very low probabilities.

Figure 2. ROC curves ‘Patient characteristics only’-model and ‘Patient characteristics and TVS’-model.
Discrimination

Figure 2 shows ROC curves for the two models and for TVS alone in the two validation datasets. In both the Dutch and the Swedish databases, discrimination as reflected by the AUC for the patient characteristics and TVS model (respectively 0.89 (95% CI 0.86-0.92) and 0.89 (95% CI 0.86-0.91) was higher compared to the AUC for the ‘patients characteristics’-model (0.71 (95% CI 0.65-0.76) in the Dutch database and 0.69 (95% CI 0.64-0.73) in the Swedish database). For TVS-only, the AUC was 0.87 (95% CI 0.83-0.90) in the Dutch Database and 0.90 (95% CI 0.88-0.93) in the Swedish database.

Comparison with current practice

In the original publication on model development, model-based strategies were compared to the currently applied strategy of endometrial thickness measurement with a cut-off value of five mm, to reflect clinical practice in the Netherlands at that time. Because the AUC’s in the ‘patient characteristics model’ and in the ‘patient characteristics and TVS’ model were calculated with the predicted probability as a continuous variable, we re-calculated the AUC for TVS-only with endometrial thickness as a continuous variable instead of a dichotomized variable (with a fixed cut-off).

The AUC for TVS-only was 0.87 (95% CI 0.83-0.90) in the Dutch and 0.90 (95% CI 0.88-0.93) in the Swedish database. The curves (figure 2) are almost identical to the curves of the ‘patient characteristics and TVS’ model.

The ‘patient characteristics’ model has a much lower AUC than TVS alone in both databases. Nevertheless, this ‘patient characteristics’ model could be used in situations where an ultrasound is not (immediately) available in the physicians’ office.

DISCUSSION

We assessed the external validity and generalizability of two previously developed clinical prediction models to estimate the probability of endometrial cancer in women presenting with PMB in two independent prospective cohorts. In this external validation study, we demonstrated that the diagnostic performance of the models was similar to the diagnostic performance of the original model in internal validation. The diagnostic performance of the ‘patient characteristics and TVS model’ however, is comparable to endometrial thickness measurement by TVS. The ‘patient characteristics’ model is able to select women with a low risk of endometrial cancer, who can be reassured without further invasive testing. In the original publication, a predicted risk lower than 4% was equal to a negative predictive value of 99%. This means that in the original model, calibration was not optimal in the low risk range; the predicted risk was higher than the observed proportion of endometrial cancer. In clinical practice, a woman can be reassured without TVS and without further testing if she is
aged 50 years or younger and has up to one additional risk factor. In a primary care setting or in a health care setting where TVS cannot be performed immediately and has to be ordered separately and be performed by a radiologist, this model is able to select women that can be reassured without TVS and without further testing.

Different calibration was observed for the models in the two different databases. In the Swedish database the predicted probability of having endometrial carcinoma is underestimated in the ‘patient characteristics and TVS’ model, except in women with a very low risk for having endometrial cancer. This difference may be due to a different composition of the two validation databases. Women in the Swedish database had lower BMI, were older, had thicker endometrium and more women had endometrial cancer.

Strengths of our current study are the external validation of the model using data from a different region within the Netherlands as well as data from another European country. The diagnostic performance of the models is comparable to the diagnostic performance in the development database, although patient groups are significantly different. Another strength is that we performed a comparison with current clinical practice: measurement of endometrial thickness in all women with PMB.

The use of the model, adding patient characteristics to endometrial thickness, does not seem to improve the efficiency of the diagnostic workup over TVS alone. With a statistical approach, we hoped to improve the diagnostic work-up for women with PMB by individualizing the strategy. In the publication on model development we concluded that accounting for the characteristics of the women could increase the efficiency. With a more clinical approach in this validation study, we find no added value of the ‘patient characteristics and TVS’ model over the use of ‘TVS-only’.

One of the limitations of this study is missing data. Multiple imputation was used to handle these missing data. Multiple imputation, even with a relatively large amount of missing data, gives a more precise and valid measure of association for variables with missing values than complete case analysis. Another limitation is the fact that in the two validation databases a different cut-off value for endometrial thickness was used. In the Dutch database, there was one patient with an endometrial thickness between 4.0 and 4.5 mm diagnosed with endometrial cancer. All women with an endometrial thickness ≤ 4.4 mm from the Swedish database were linked to the national cancer registry. None of these women were diagnosed with endometrial cancer during follow-up. Therefore we think that the different cut-offs used in two databases did not influence our results.

Several prediction models have been published to estimate the risk of endometrial cancer in women with postmenopausal bleeding. Opolskiene et al developed a prediction model for women with an endometrial thickness ≥ 4.5 mm focusing on the combination of clinical and ultrasound characteristics: endometrial thickness and power Doppler examination of the endometrium. Burbos et al. developed two models based on patient characteristics
with and without endometrial thickness, similar to the models we published in 2007.\textsuperscript{18, 30, 31} None of these models were externally validated.\textsuperscript{29} External validation, assessing the validity and generalizability of a model is an essential step before a model can be implemented in practice.\textsuperscript{25, 32} Our study is the first to describe external validation of a model estimating the risk of endometrial carcinoma in women with PMB.

**Conclusions**

This study shows that after external validation in two independent datasets, the existing multivariable models maintain their diagnostic performance and are able to distinguish between women with low or high risk of endometrial cancer in women with postmenopausal bleeding not using HRT. The ‘patient characteristic and TVS’ model however, offers no diagnostic advantage over the measurement of endometrial thickness alone. The ‘patient characteristics’ model is able to select women with a low risk of endometrial cancer, who can be reassured without further testing. This is especially useful in a setting where TVS is not (directly) available.
REFERENCES