The diagnostic work-up of women with postmenopausal bleeding
van Hanegem, N.

Citation for published version (APA):
van Hanegem, N. (2015). The diagnostic work-up of women with postmenopausal bleeding

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.
Chapter 4

External validation of a mathematical model to estimate the probability of endometrial cancer of women with postmenopausal bleeding

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

Submitted
Abstract

Objective To externally validate two previously developed prediction models that estimate the probability of endometrial cancer in women with postmenopausal bleeding.

Design and setting We performed an external validation study of two previously developed prediction models in two independent datasets of consecutive women not using hormone replacement therapy with a first episode of postmenopausal bleeding.

Population We studied women with postmenopausal bleeding. One dataset (559 women) was prospectively collected in three general hospitals in the Netherlands including, the other dataset (433 women) was prospectively collected in a university hospital in Sweden.

Methods We retrospectively evaluated two models that predict endometrial cancer in the two validation databases. We then evaluated three diagnostic strategies, a ‘patient characteristics’ rule, based on characteristics of the women without transvaginal ultrasound, a ‘sequential’ rule, i.e. ultrasound in case the probability for cancer exceeded 4% based on characteristics, and subsequent histological analyses when the endometrial thickness exceeded 4 mm and an ‘integrated’ rule with a probability estimate based on both characteristics of the women and ultrasound results and endometrial sampling when the probability of cancer exceeded 4%.

Main outcome measures We studied the performance of the models in terms of discrimination and calibration. We then calculated the number of carcinomas detected and missed using the three different strategies, as well as the number of ultrasounds and invasive procedures performed with the three different strategies.

Results In both the Dutch and the Swedish databases, the two models showed good performance in terms of discrimination and calibration. The three strategies, based on these two models, all detected all women with endometrial (pre) cancer. Applying the ‘integrated’ or ‘sequential’ strategy would, compared to current practice (ultrasound only), leads to a 3 to 6 % decrease in the number of women in need for further invasive testing.
**Conclusions** We found that two models for endometrial cancer maintained their diagnostic performance in two independent validation databases. The use of a 'patient characteristic and ultrasound' model in a sequential or integrated strategy, could slightly reduce the number of invasive procedures without loss in detection of endometrial cancer.
Introduction

Postmenopausal bleeding (PMB) is a common complaint in postmenopausal women and in about 10% of women endometrial cancer is the underlying cause of PMB.\(^1\)

In the 1990s endometrial thickness measurement with transvaginal ultrasonography (TVS) was introduced as a test to distinguish between women with a low and a high risk of endometrial cancer.\(^2\) The cut-off point for a thin endometrium, and thus a low risk of endometrial cancer, varies in different guidelines between three and five millimetres.\(^3\) Patients with a thin endometrium can be reassured as their post-test probability of endometrial cancer is lower than one per cent, which is a worldwide-accepted threshold for patient reassurance.\(^1,4,5\) The post-test probability depends not only on the endometrial thickness, 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, anticoagulants use and parity.\(^6\)\(^-\)\(^12\)

Several studies have described the prevalence of these characteristics and developed different prediction models to estimate the individual chance of having endometrial cancer.\(^13\) However, none of the existing prediction models have yet been externally validated, which is necessary for successful implementation.\(^14\) All models were internally validated in their development database and two models showed the best performance.\(^15,16\) Opolskiene et al concluded that their model excludes endometrial cancer reasonably well when power Doppler is added, but because Doppler is not commonly used in daily practice, we decided to validate two multivariable models without Doppler, described by Opmeer et al\(^15,16\) These two models had been internally validated in their development database.\(^13\)

The aim of the present study was to externally validate the diagnostic performance of these two models and estimate the clinical consequences of the three management strategies suggested in this article by retrospectively applying the models on two independently prospectively collected databases of women with PMB.

Methods

The multivariable models and management strategies

Opmeer et al developed two multivariable logistic regression models for the
prediction of endometrial cancer in women with postmenopausal bleeding based on the following patient characteristics: age, time since menopause, body mass index (BMI), diabetes, parity, hypertension, use of anticoagulants, history of cancer and dysfunction of the thyroid gland. The first statistical model, based on characteristics of the women only, is referred to as the ‘patient characteristics model’. The second model is an extension of the first model in which characteristics of the women are combined with the measurement of endometrial thickness by TVS and is referred to as the ‘patient characteristics and TVS model’. Further statistical details on the development of the model can be found in the Appendix of the original article.16

Study population used for external validation
The two models and three strategies were externally validated by retrospectively applying them on two prospectively collected databases:

- Dutch database: This database studying 559 women included all women presenting with postmenopausal bleeding between January 2009 and April 2011 (36% of women presented at the TweeSteden Hospital, 18% at the Maxima Medical Centre in Veldhoven and 46% at the St. Antonius Hospital in Nieuwegein, the Netherlands). This database was not primarily established to validate the two mathematical models of Opmeer et al. Menopause was defined as at least one year of amenorrhea, after the age of forty. If there were doubts about menopausal status the status was established by testing of the follicle-stimulating hormone (FSH). 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 endometrial thickness exceeded four millimetres, endometrial sampling using the Pipelle® (Laboratoire CCD, Paris, France) was performed. In case of a failed endometrial sampling hysteroscopy with directed biopsy was performed. Failure was classified as either a technical failure or as an endometrial sampling in which the amount of tissue was insufficient for a reliable diagnosis. All women were instructed to contact the hospital if recurrence of bleeding occurred.

- Swedish database: This database includes all women presenting with PMB at the Skåne University Hospital in 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 measured by TVS. If endometrial thickness exceeded 4.4 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® (Medscan AB, Malmö, Sweden) was taken. If there were focal lesions at SIS or if SIS failed, hysteroscopy with resection of focal lesions (if present) and supplementary dilatation and curettage were performed. Women in whom endometrial sampling was not performed because of endometrium ≤ 4.4 mm 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, who did not have endometrial sampling without recurrent bleeding, were considered negative for endometrial cancer. In case of recurrent bleeding, hysteroscopy was performed. 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 ≤ 4.4 mm (and therefore without a histological diagnosis of the endometrium) were matched with the regional cancer register to ascertain that none of these women were diagnosed with endometrial cancer after inclusion in the study. Indeed, no woman with endometrial thickness ≤ 4.4 mm was found to have endometrial cancer and all were classified as not having endometrial cancer.

Women were excluded from our statistical analysis if the endometrial thickness was not measurable. The Swedish database does not include women with fluid in the uterine cavity. Precancer, defined as any form of hyperplasia with atypia, and cancer in the histology specimen were classified as ‘endometrial cancer’. All other histological diagnoses were classified as benign.

**Statistical analysis of patient characteristics**

We compared patient characteristics between the two databases with the chi-square test. Continuous variables were tested for normal distribution, and the independent
T-test or Mann–Whitney U-test for univariate analysis was used to compare means or medians. For analysis, the Statistical Package for the Social Sciences (IBM Corp, Armonk, NY, USA) version 20.0 was utilized. Statistical significance was set at p < 0.05.

**Imputation of missing values**

In our validation study, we performed multiple imputations for missing data elements, with separate imputation rounds for each of the two databases. 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 pooled estimates and standard errors reflect missing data uncertainty.\(^\text{17-19}\)

**External validation**

We retrospectively applied the two models developed by Opmeer (patient characteristics only, patient characteristics and TVS) to the women in the Dutch and Swedish databases. We assessed the performance of the models by examining calibration (agreement between predicted risks and observed frequencies of endometrial cancer) and discriminative performance (the ability of the models to distinguish between women with and without endometrial cancer). To assess calibration for the two models, 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.\(^\text{20}\) Calibration is considered perfect if the intercept is 0 and the calibration slope is 1.\(^\text{21,22}\) Calibration is relevant to evaluate the accuracy of the risk estimates provided by the models (do patients with predicted risk of 25% indeed have a risk of 1 in 4 of having endometrial cancer), but in clinical practice high performance in terms of identified and missed cases at a certain threshold will be required. Calibration analyses were performed using R version 15.2.1.

Discriminative performance of the two models was assessed by calculating the area under the receiver operator characteristic curve (AUC). AUCs reflect the overall discriminative taking into account the full spectrum of predicted probabilities. As such, they are informative from a statistical perspective, but a model with lower AUC may show superior clinical performance at a particular threshold as compared to a model with higher AUC.
Based on these two statistical models, three different diagnostic strategies were explored:

1. The ‘patient characteristics’ rule, i.e. probability estimates based on characteristics of the women, and invasive diagnostics in case the probability of (pre) cancer exceeded 4%. In this rule, TVS was not performed.
2. The ‘sequential’ rule, i.e. probability estimates based on characteristics of the women, with TVS in case the probability for cancer exceeded 4%, and subsequent histological analyses when the endometrial thickness exceeds 4 mm.
3. The ‘integrated’ rule, i.e. TVS in all women, with a probability estimate based on both characteristics of the women and TVS results, completed by endometrial sampling when the probability of cancer exceeded 4%.

To estimate the clinical consequences of applying these three different management strategies proposed by Opmeer et al on the validation datasets, we calculated for each of the three strategies, the percentage of women in whom TVS would be performed, the percentage of women in whom an invasive procedure would be performed to obtain material for histology, the number of endometrial (pre) cancers identified by the different strategies and specificity (the amount of patients without cancer who fell below the threshold of the specific strategy). We compared all these clinical consequences with the current strategy used in clinical practice: selecting women for further diagnostic work-up by the measurement of endometrial thickness (TVS only). We performed all analyses for the external validation in ‘R’, version 2.15.0 (2012).

Results

Patient characteristics

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 prevalence 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 Swedish women used anticoagulant therapy. For both validation databases, efforts were made to collect data on endometrial cancer in the women with an endometrial thickness below the applied threshold and in none of these women endometrial cancer was diagnosed.

Table 1. Patient characteristics and missing values in validation databases

<table>
<thead>
<tr>
<th></th>
<th>Swedish database</th>
<th>Missing; n (%)</th>
<th>Dutch database</th>
<th>Missing; n (%)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>433</td>
<td>559</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age, years; mean (SD)</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>Diabetes mellitus; n (%)</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>Hypertension; n (%)</td>
<td>176 (40.6)</td>
<td>0</td>
<td>196 (35.1)</td>
<td>0</td>
<td>0.07</td>
</tr>
<tr>
<td>Anticoagulants; n (%)</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/m²; mean (SD)</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>Time since menopause, years; median (IQR)</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; n (%)</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>Endometrial thickness, mm; median (IQR)</td>
<td>6.0 (3.2-13.0)</td>
<td>0</td>
<td>5.7 (2.5-10.0)</td>
<td>0</td>
<td>0.02</td>
</tr>
<tr>
<td>Endometrial (pre-) cancer; n (%)</td>
<td>65 (15.0)</td>
<td>0</td>
<td>57 (10.2)</td>
<td>0</td>
<td>0.02</td>
</tr>
<tr>
<td>Atypical hyperplasia</td>
<td>3 (0.69)</td>
<td>7 (1.3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Endometrial cancer</td>
<td>62 (14.3)</td>
<td>50 (8.9)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

BMI, body mass index; NA, not applicable.

Calibration

Plots that express the calibration of the two models developed by Opmeer ('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 was better for the 'patient characteristics and TVS' model than for the 'patient characteristics' model. The predicted probabilities of cancer when using the 'patient characteristics and TVS' model were close to the observed frequency of cancer over the whole range of predicted risks. For the lowest risk group, calibration was consistently good across the different models and databases, and some major over and underestimated risk in patients at increased risk for endometrial cancer.
Figure 1. Calibration plots for the two prediction models

A. ‘Patient Characteristics’ model Dutch validation database

B. ‘Patient Characteristics and TVS’ model Dutch validation database
External validation of a prediction model

In the Swedish database, the calibration slope performed better for the ‘patient characteristics’ model than for the ‘patient characteristics and TVS’ model. For the ‘patient characteristics’ model, predicted probabilities were close to the observed frequency over the whole range of predicted risks. The ‘patient characteristics and TVS’ model underestimated the probability of endometrial cancer over almost the whole range of probabilities.
Discriminative performance of the prediction models

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

Clinical consequence of the three strategies

The estimated clinical consequences of applying the three strategies are reported in Tables 2 and 3. With all three strategies, all cases of endometrial (pre) cancers would be detected in both databases. This is under the assumption that among women with a thin endometrium (below the threshold of 4 or 4.4 mm) was diagnosed with endometrial cancer. This means that with the ‘patient characteristics’ strategy you could skip the measurement of endometrial thickness safely, but then you would have to perform an invasive procedure to get histology in 93% of women in both databases, compared to only 61–63% (respectively in the Dutch and Swedish database) when patients would be selected based on the measurement of endometrial thickness (which is current clinical practice).

When using the patient characteristics in a ‘sequential’ strategy, one could save 7% of women an ultrasound and these could be reassured. In the remaining group of women endometrial thickness has to be measured and then 57–58% would have to undergo an invasive procedure. Thus, this strategy would save 7% of women an ultrasound and in 3–6% less women an invasive procedure has to be done compared to the current clinical practice: TVS only.

When using the patient characteristics in an ‘integrated’ strategy, all women would need a TVS and the amount of women that would need further invasive procedures to retrieve histology would be the same as in the ‘sequential’ strategy.
External validation of a prediction model

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

A. Dutch validation database
- ‘Patient characteristics’-model, AUC: 0.71 (95% CI: 0.65 to 0.76)
- ‘Patient characteristics and TVS’-model, AUC: 0.89 (95% CI: 0.86 to 0.92)
- TVS only, AUC: 0.87 (95% CI: 0.83 to 0.90)

B. Swedish validation database
- ‘Patient characteristics’-model, AUC: 0.69 (95% CI: 0.64 to 0.73)
- ‘Patient characteristics and TVS’-model, AUC: 0.89 (95% CI: 0.86 to 0.91)
- TVS only, AUC: 0.90 (95% CI: 0.88 to 0.93)
### Table 2. Clinical consequences of strategies, Dutch database

<table>
<thead>
<tr>
<th>Decision strategy</th>
<th>Description</th>
<th>TVS N (%)</th>
<th>Invasive procedures N (%)</th>
<th>(Pre) cancers detected N (%)</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference strategy</td>
<td>TVS only (current practice) &gt; 4mm</td>
<td>559 (100%)</td>
<td>352 (63.0%)</td>
<td>57 (100%)</td>
<td>0.41</td>
</tr>
<tr>
<td>'Patient characteristics' strategy</td>
<td>Histological analysis if probability based on characteristics of the women exceeds 4% (no TVS)</td>
<td>0 (no TVS)</td>
<td>521 (93.2%)</td>
<td>57 (100%)</td>
<td>0.08</td>
</tr>
<tr>
<td>'Sequential' strategy</td>
<td>Decision for TVS based on probability of cancer calculated from characteristics of the women, TVS only performed when probability exceeds 4%, histological analysis if TVS &gt; 4mm</td>
<td>521 (93%)</td>
<td>320 (57.2%)</td>
<td>57 (100%)</td>
<td>0.48</td>
</tr>
<tr>
<td>'Integrated' strategy</td>
<td>TVS in all women, decision for histological analysis if probability for endometrial cancer calculated from characteristics of the women and TVS model exceeds 4%</td>
<td>559 (100%)</td>
<td>316 (56.5%)</td>
<td>57 (100%)</td>
<td>0.48</td>
</tr>
</tbody>
</table>

### Table 3. Clinical consequences of strategies, Swedish database

<table>
<thead>
<tr>
<th>Decision strategy</th>
<th>Description</th>
<th>TVS N (%)</th>
<th>Invasive procedures N (%)</th>
<th>(Pre) cancers detected N (%)</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference strategy</td>
<td>TVS only (current practice) &gt; 4mm</td>
<td>433 (100%)</td>
<td>265 (61.2%)</td>
<td>65 (100%)</td>
<td>0.46</td>
</tr>
<tr>
<td>'Patient characteristics' strategy</td>
<td>Histological analysis if probability based on characteristics of the women exceeds 4% (no TVS)</td>
<td>0 (no TVS)</td>
<td>401 (92.6%)</td>
<td>65 (100%)</td>
<td>0.09</td>
</tr>
<tr>
<td>'Sequential' strategy</td>
<td>Decision for TVS based on probability of cancer calculated from characteristics of the women, TVS only performed when probability exceeds 4%, histological analysis if TVS &gt; 4mm</td>
<td>401 (92.6%)</td>
<td>251 (58.0%)</td>
<td>65 (100%)</td>
<td>0.49</td>
</tr>
<tr>
<td>'Integrated' strategy</td>
<td>TVS in all women, decision for histological analysis if probability for endometrial cancer calculated from characteristics of the women and TVS model exceeds 4%</td>
<td>433 (100%)</td>
<td>248 (57.2%)</td>
<td>65 (100%)</td>
<td>0.50</td>
</tr>
</tbody>
</table>
Discussion

In this external validation study, we demonstrate that the diagnostic performance (i.e. Discrimination and calibration) in external validation of the two models is similar to the discriminative performance of the models in internal validation in the development database. The model with the best discriminative performance in this external validation is the patient ‘characteristics and TVS’ model. However, we found that the discriminative performance of the ‘patient characteristics and TVS’ model is similar to that of endometrial thickness measurements with TVS (TVS-only), i.e. adding clinical information to endometrial thickness measurement does not significantly improve the ability to discriminate between benign and malignant endometrium.

Applying a strategy basing the decision to perform further invasive diagnostics on an individual risk calculated with ‘patient characteristics only’ would be safe. This means, no endometrial (pre) cancer that would have been detected by selecting women with TVS would be missed if these women would be selected based on patient characteristics only. However, you would need to perform invasive diagnostics in 93% of women, compared to only 61-63% (respectively in the Dutch and Swedish database) when patients would be selected based on TVS.

An important strength of our study is the external validation of the models using data from a different region within the Netherlands as well as data from another European country. External validation, assessing the validity and generalizability of a model is an essential step before a model can be implemented in practice. To our knowledge this is the first study to describe external validation of a prediction model estimating the risk of endometrial cancer in women with PMB. As the TweeSteden Hospital also participated in the development study of the two prediction models by Opmeer et al, the population used for external validation has a minor overlap with the development population, yet with completely separate samples (different women in the development and validation sample).

As many data were collected as part of clinical practice, not all information was available for all women. Multiple imputation was used to deal with 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. Generally, dropping cases with missing values (complete case analysis) yields biased results, and the discriminative ability of a multivariable model is reduced when cases with missing values are excluded from analysis.
Another limitation is the fact that partial verification was performed in both databases. In women with an endometrial thickness below the applied threshold no histological assessment was performed. This is in agreement with clinical policy since more than 15 years in both Sweden and the Netherlands, and for practical and ethical reasons we have not included this assessment for research purposes only. For both validation databases however, efforts were made to collect information on these women by assessing patient charts and in the Swedish database matching the patients with the regional cancer registry. Evidently, there remains some uncertainty whether indeed no endometrial cancers were missed in this group, but with our approach we minimised this risk by our follow-up efforts.

Several prediction models have been published to estimate the risk of endometrial cancer in women with postmenopausal bleeding. Opolskiene et al developed four prediction models including clinical and ultrasound information for women with endometrial thickness $\geq$ 4.5 mm. The first model is based on patient characteristics and in the development database an AUC of 0.74 was found. The second model is based on patient characteristics and endometrial thickness as measured with TVS, with an AUC of 0.82. The last two models are based on patient characteristics combined with sonographic endometrial thickness and two different Doppler characteristics, with AUC of 0.89 and 0.91 respectively. The authors concluded that the models are fairly good in excluding endometrial cancer when power Doppler is added. Burbos et al developed two models based on patient characteristics with (AUC 0.77) and without endometrial thickness (AUC 0.73). The authors concluded that the model based on patient characteristics has a reasonable discriminatory ability and the model based on patient characteristics and endometrial thickness has a fair accuracy in separating women without cancer from women with cancer. These findings are similar to the findings of internal validation of the models by Opmeer et al. None of these models have yet been externally validated.

Before a prediction model can be implemented in clinical practice, external validation is essential. In this external validation we found that the ‘patient characteristics and TVS’ model shows good discriminative performance (AUC) and a reasonable performance on calibration, however it is comparable to the use of TVS-only. All three strategies based on the ‘patient characteristics and TVS’ model could be safely implemented in daily practice, i.e. without missing any additional (pre) cancers compared to TVS-only, which is the current daily practice. To choose which strategy is used best in clinical practice, one could focus on the availability and use of different diagnostic tests. In situations were no ultrasound is available, women...
could be selected based on their characteristics, however with a very low specificity which means that many women need to undergo further invasive testing with a small chance of diagnosing a (pre) cancer.

This study shows that after external validation in two independent datasets, the two multivariable models maintain their diagnostic performance and are able to select women with PMB not using HRT with a low risk of having endometrial cancer. By using the ‘patient characteristic and TVS’ model in a sequential or integrated strategy, the number of women that need further invasive procedures to obtain material for histological assessment can be decreased only a little, compared to the current diagnostic pathway in which patient characteristics are not taken into account. Therefore we think that these models are moderately useful in current daily practice. In the Netherlands, TVS is easy accessible in daily practice. A model based on patient characteristics would only be useful if a larger amount of invasive procedures could be decreased. Future research should focus on adjusting the model with different thresholds or subgroups of patients. Furthermore, as Doppler will be used more and more in clinical practice it would be worth it to externally validate a model which uses Doppler features.

**Conclusions**

The two models for endometrial cancer based on patient characteristics and TVS, maintained their diagnostic performance in two independent validation databases. The use of a ‘patient characteristic and TVS’ model in a sequential or integrated strategy, could slightly reduce the number of invasive procedures, compared to the current diagnostic pathway in which patient characteristics are not taken into account, without loss in detection of endometrial cancer. 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.
Chapter 4

References


External validation of a prediction model

