



Development and external validation of new ultrasound-based mathematical models for preoperative prediction of high-risk endometrial cancer

C. VAN HOLSBEKE*†, L. AMEYE‡, A. C. TESTA§, F. MASCILINI§, P. LINDQVIST¶, D. FISCHEROVA**, F. FRÜHAUF**, S. FRANSIS††, E. DE JONGE†, D. TIMMERMAN* and E. EPSTEIN¶

*Department of Obstetrics and Gynaecology, University Hospitals Leuven, Leuven, Belgium; †Department of Obstetrics and Gynaecology, Ziekenhuis Oost-Limburg, Genk, Belgium; ‡Department of Development and Regeneration, KU Leuven - University of Leuven, Leuven, Belgium; §Istituto di Clinica Ostetrica e Ginecologica, Università Cattolica del Sacro Cuore, Roma, Italy; ¶Department of Obstetrics and Gynecology, Karolinska University Hospital, Stockholm, Sweden; **Gynecological Oncology Center, Department of Obstetrics and Gynecology, First Faculty of Medicine and General University Hospital, Charles University, Prague, Czech Republic; ††Department of Pathology, Ziekenhuis Oost-Limburg, Genk, Belgium

KEYWORDS: endometrial cancer; high risk; mathematical models; prediction

ABSTRACT

Objectives To develop and validate strategies, using new ultrasound-based mathematical models, for the prediction of high-risk endometrial cancer and compare them with strategies using previously developed models or the use of preoperative grading only.

Methods Women with endometrial cancer were prospectively examined using two-dimensional (2D) and three-dimensional (3D) gray-scale and color Doppler ultrasound imaging. More than 25 ultrasound, demographic and histological variables were analyzed. Two logistic regression models were developed: one 'objective' model using mainly objective variables; and one 'subjective' model including subjective variables (i.e. subjective impression of myometrial and cervical invasion, preoperative grade and demographic variables). The following strategies were validated: a one-step strategy using only preoperative grading and two-step strategies using preoperative grading as the first step and one of the new models, subjective assessment or previously developed models as a second step.

Results One-hundred and twenty-five patients were included in the development set and 211 were included in the validation set. The 'objective' model retained preoperative grade and minimal tumor-free myometrium as variables. The 'subjective' model retained preoperative grade and subjective assessment of myometrial invasion. On external validation, the performance of the new

models was similar to that on the development set. Sensitivity for the two-step strategy with the 'objective' model was 78% (95% CI, 69–84%) at a cut-off of 0.50, 82% (95% CI, 74–88%) for the strategy with the 'subjective' model and 83% (95% CI, 75–88%) for that with subjective assessment. Specificity was 68% (95% CI, 58–77%), 72% (95% CI, 62–80%) and 71% (95% CI, 61–79%) respectively. The two-step strategies detected up to twice as many high-risk cases as preoperative grading only. The new models had a significantly higher sensitivity than did previously developed models, at the same specificity.

Conclusion Two-step strategies with 'new' ultrasound-based models predict high-risk endometrial cancers with good accuracy and do this better than do previously developed models. Copyright © 2013 ISUOG. Published by John Wiley & Sons Ltd.

INTRODUCTION

Endometrial cancer is the most common gynecologic malignancy in developed countries, with an incidence of 12.9 per 100 000 women and a mortality rate of 2.4 per 100 000 women¹. Prognosis in the early stages of the disease (Stage I–II) depends on age, histological type and grading, depth of myometrial invasion, cervical invasion and lymph-node metastases. Most women with endometrial cancer are diagnosed at an early stage when the prognosis is excellent². However, women with

Correspondence to: Dr C. Van Holsbeke, Department of Obstetrics and Gynaecology, Z.O.L. Genk, Belgium Schiepse Bos 6, B-3600 Genk, Belgium (e-mail: caroline.van.holsbeke@skynet.be)

Accepted: 17 September 2013

high-risk endometrial cancer (high grade, deep myometrial invasion or cervical invasion) have a worse prognosis. These women are at a higher risk of recurrence and might benefit from extended surgical staging with systematic pelvic and para-aortic lymphadenectomy in order to tailor adjuvant therapy. On the other hand, it is important to avoid overtreatment, as most of these women are elderly with comorbidities. Lymphadenectomy shows no survival benefit in low-risk endometrial cancer patients and, in fact, results in increased complications and higher morbidity^{3–7}.

Thus far, preoperative risk stratification is mainly based on the use of preoperative grading (classifying Grade 3 endometrioid and the non-endometrioid histological subtypes (clear cell or serous carcinomas and carcinosarcomas) as high risk) and computed tomography scans to screen for enlarged lymph nodes. In most countries preoperative sonographic assessment of the uterus to evaluate myometrial infiltration and cervical invasion is not yet standard practice. However, several studies have assessed subjective assessment of myometrial invasion using transvaginal ultrasound, with sensitivities of 68–93% and specificities of 82–83%^{8–11}. There are fewer publications on the use of ultrasound to assess cervical invasion, but the results are generally very good^{8,10,12}. The accuracy of ultrasonography seems to be comparable with that of magnetic resonance imaging (MRI)^{13–16}. Nevertheless, at present, MRI is less accessible because it is more expensive and time consuming and, in most centers, the waiting lists are quite long. Therefore, if access to MRI is limited, priority could be given to patients with poor image quality during ultrasound as a result of obesity, fibroids, adenomyosis or the anatomic position of the uterus.

Currently, most studies that use ultrasound to predict high-risk endometrial cancer or deep myometrial invasion are based on the performance of subjective assessment or ultrasound parameters only. Our hypothesis was that mathematical models combining ultrasound data with other available preoperative information could improve the ability to identify high-risk endometrial cancer.

The primary aim of this study was to assess whether new mathematical models that include ultrasound, demographic variables and grade may improve the preoperative identification of women with high-risk endometrial cancer. The secondary aim was to validate the new models and compare them with previously published models^{17,18}.

METHODS

This was a prospective international multicenter study that included patients with endometrial cancer confirmed histologically between January 2007 and May 2009. The four centers that contributed to this study were the gynecologic ultrasound unit at the Department of Obstetrics and Gynecology, Lund University Hospital, Sweden; the Gynecologic Oncology Unit, Catholic University of the Sacred Heart, Rome, Italy; the Department of Obstetrics and Gynecology, Leuven University Hospitals, Leuven, Belgium and the Department of Obstetrics

and Gynecology of Ziekenhuis Oost-Limburg, Genk, Belgium; and the Gynecological Oncology Center and Department of Obstetrics and Gynecology, Charles University, Prague, Czech Republic. Ethical approval was obtained from the Ethics Committee of Lund University (LU-412-07) and from the local Ethics Committees of the Catholic University of the Sacred Heart (Rome, Italy), University Hospitals Leuven (Leuven, Belgium) and Ziekenhuis Oost-Limburg (Genk, Belgium).

We included patients with a histologically proven diagnosis of endometrial cancer obtained by pipelle, dilatation and curettage (D&C) or hysteroscopy. All patients underwent a standardized ultrasound examination by an experienced ultrasound examiner (C.V.H., E.E., A.C.T. or D.F.) following a strict protocol. The two-dimensional (2D) and three-dimensional (3D) gray-scale and vascular pattern criteria were defined at a consensus reading session including all examiners using the International Endometrial Tumor Analysis (IETA) terminology¹⁹. The objective ultrasound variables that were assessed during real-time 2D examination included: tumor size in three perpendicular diameters; minimal tumor-free margin (the minimal thickness of the tumor-free myometrium, i.e. the area at which the tumor is invading most (measured perpendicular to the serosa)); the tumor/uterine anteroposterior (AP) diameter ratio; tumor echogenicity; the endometrial/myometrial border (regular/irregular); the presence and number of fibroids; the distance from the outer cervical os to the caudal border of the lesion; and vascularization characteristics. The amount of tumor vascularization was classified subjectively using a 'color score' (absent (= 1), minimal (= 2), moderate (= 3) or high (= 4)), as introduced by the International Ovarian Tumor Analysis (IOTA) group²⁰. From the stored 3D volumes we calculated the tumor volume, the uterine volume and the tumor volume/uterine volume ratio, and we assessed vascular morphology using GE 4D View (GE Healthcare, Little Chalfont, UK) or Philips QLAB (Koninklijke Philips N.V., Amsterdam, The Netherlands), using 30° rotational steps. The examiners also reported on their subjective assessment of myometrial and cervical infiltration. Subjective assessment is the personal impression of the ultrasound examiner of depth of invasion and cervical invasion whilst scanning the uterus. Patients were classified as high risk when the ultrasound examiner believed that the tumor invaded one half or more of the myometrium or when cervical stromal invasion was suspected.

The ultrasound equipment used in Lund was a GE Voluson E8 ultrasound system with a RIC5-9 transducer or a Philips IU22 Ultrasound system with a 3D9-3v transducer. In the hospitals in Rome, Leuven and Genk, the GE Voluson E8 with a RIC5-9 transducer was used. Static images with all measurements, videoclips with and without power Doppler, and 3D volumes with and without power Doppler were collected for every patient.

All women underwent an extrafascial abdominal hysterectomy with bilateral salpingoophorectomy and systematic pelvic and para-aortic lymphadenectomy, if appropriate, according to the local protocols. The

International Federation of Gynecology and Obstetrics (FIGO) stage that was assigned to each case was based on final histology. In each center, the surgical specimens were examined by a dedicated pathologist with substantial experience in gynecologic oncology using a predetermined protocol. The histopathological variables that were assessed were: histological subtype (endometrioid/non-endometrioid; based on preoperative biopsy and on hysterectomy); grade of differentiation (based on preoperative biopsy and on hysterectomy); tumor size, macroscopic growth pattern and macroscopic estimation of infiltration; number of fibroids; microscopic infiltration; minimal tumor-free myometrium; cervical invasion; extension to parametria; and FIGO classification. The tumors were classified according to the criteria recommended by the World Health Association (WHO) and the pathological tumor stage according to the criteria recommended by the International Union Against Cancer (TNM Classification of malignant tumors)²¹. The FIGO 2009 staging criteria were used²². Only epithelial malignant tumors and carcinosarcomas (malignant mesodermal mixed tumors (MMMTs)) were included. Endometrial adenocarcinoma was classified into three grades (Grade 1, well differentiated; Grade 2, moderately differentiated; and Grade 3, poorly differentiated). Other histological subtypes, such as serous/seropapillary, clear cell and carcinosarcoma, were classified as Grade 3. The final allocation of high-risk status of a tumor was based on the microscopic findings in the hysterectomy specimen (i.e. Stage Ib or more, or Grade 3/non-endometrioid histotype (serous, clear cell or carcinosarcoma)).

The prediction of high-risk cancer was performed using either the preoperative grading only (i.e. the 'one-step strategy') or a 'two-step strategy' (Figure 1). In the first phase of the study, all patients with a poorly differentiated carcinoma or a non-endometrioid cancer (serous, clear cell or carcinosarcoma) on preoperative biopsy were automatically classified as high-risk cancers. For the remaining patients (Grade 1 and Grade 2) we created mathematical models to predict high-risk cancer, ultimately adding the results from both groups. In the second phase we externally validated the new models used in the two-step strategy and compared their performance with previously published strategies.

The gold standard was the FIGO classification that was assigned after final histopathological analysis of the specimen.

For the development of a new two-step model we used 125 women examined in Lund, Leuven/Genk and Rome. Among these, 94 had a Grade 1 or a Grade 2 tumor based on preoperative histology. The external validation was based on 211 new consecutive cases examined in Prague that had not contributed to the development set; 163 of these cases had a Grade 3 tumor on preoperative histology and were used for model validation.

We developed two new logistic regression models. The 'objective model' considered all the variables listed in the univariate analysis and subjective assessment of cervical invasion as possible predictors. The 'subjective

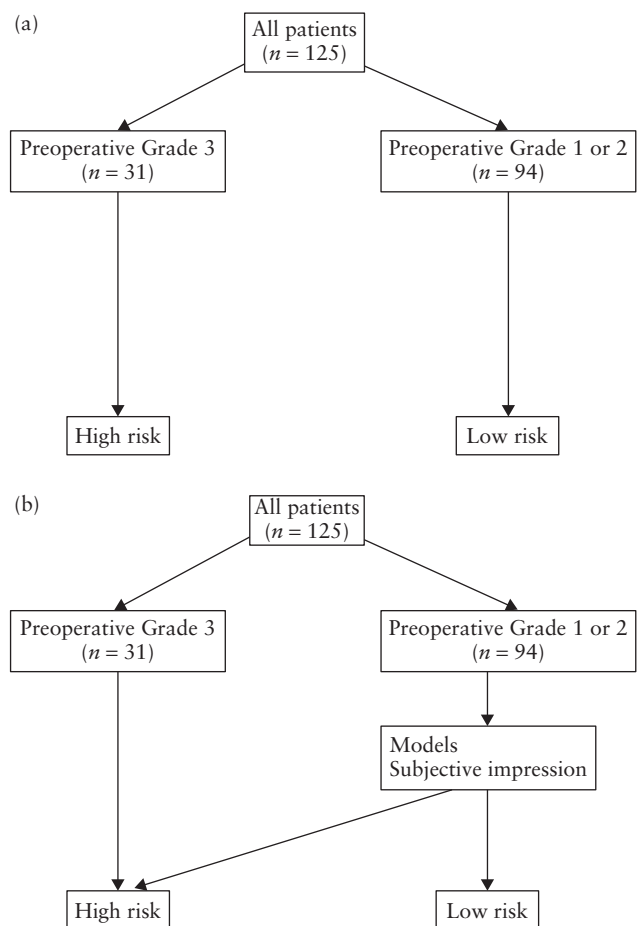


Figure 1 Flow charts for the prediction of high-risk endometrial cancer using one-step (a) and two-step (b) strategies.

model' considered subjective variables (i.e. subjective impression of myometrial and cervical stromal invasion, preoperative grading and all demographic variables) as possible predictors.

Old models in the validation included:

- The Karlsson criteria. Karlsson defined the following criteria to predict deep myometrial infiltration; the maximal AP thickness of the endometrial lesion measured in the sagittal plane (d1) divided by the AP uterine diameter (d2) (Figure 2). More than 50% myometrial infiltration was defined as $d2/d1 \geq 50\%$; less than 50% myometrial infiltration was defined as $d2/d1 < 50\%$ ¹⁰. Cervical infiltration was suspected when there was no clear demarcation of the endometrial lesion toward the cervical canal¹⁷.
- The logistic regression model of De Smet *et al.* This mathematical model was developed to predict deep myometrial invasion (> 50% infiltration), and variables selected included preoperative grade, number of fibroids, endometrial thickness and endometrial volume (calculated from three measurements of the endometrium in two perpendicular planes)¹⁸.

Statistical analysis

All ultrasound and demographic variables were first assessed in a univariate analysis. Differences between high

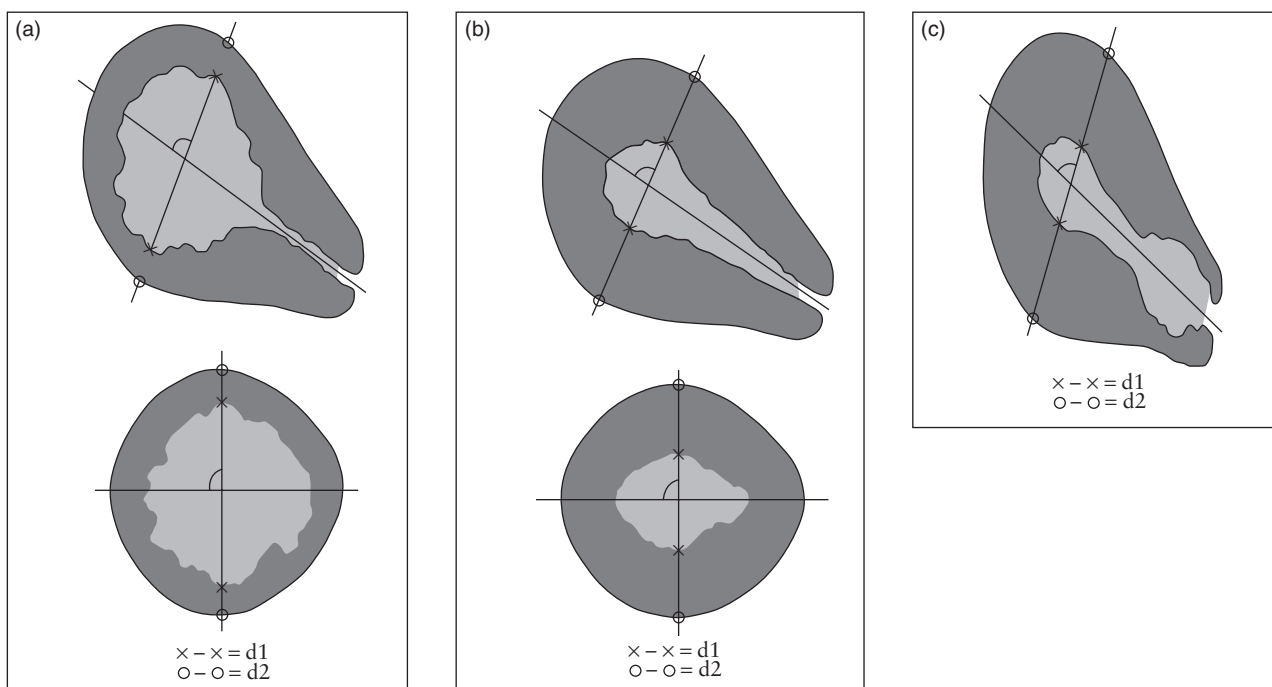


Figure 2 Schematic representation, according to the criteria of Karlsson¹⁷, to demonstrate: (a) deep myometrial infiltration ($d2/d1 \geq 50\%$), myometrial infiltration of less than 50% ($d2/d1 < 50\%$) and (c) cervical infiltration.

risk and low risk in continuous variables were assessed using a *t*-test and the Mann–Whitney *U*-test; differences in categorical variables were assessed using the chi-square test and Fisher's exact test.

As described previously, we decided to create two new mathematical models: one containing mainly objective measurements, 'the objective model', and the other also including subjective assessment of cervical and myometrial invasion. By the use of stepwise logistic regression we selected variables for the models. Only variables with a univariate $P < 0.2$ were considered as possible predictors in the multivariate model.

The different prediction strategies were tested at two different cut-off levels: 0.40 and 0.50. The cut-off level of 0.50 gave the best sensitivity for an acceptable level of specificity. The cut-off level of 0.40 was tested to evaluate the level of specificity when increasing sensitivity. The cut-off levels represent values for the mathematical model and are not related to absolute risk. To compare differences in sensitivity and specificity between the different strategies, the McNemar test was used. All statistical analyses were performed with SAS 9.3 (SAS Institute Inc., Cary, NC, USA). All *P*-values were two-sided and $P < 0.05$ was considered significant.

RESULTS

Of the 144 patients initially included in this study, 19 were excluded because no information on preoperative grade was available. The demographic and clinical characteristics for women with high- and low-risk cancer from development and validation sets, according to final histology, are presented in Table 1. Women with high-risk

cancer were significantly older than were women with low-risk cancer.

The preoperative histology, final histology and final staging of development and validation sets are presented in Table 2. Around half of the tumors were Stage IA, in both development and validation sets. On the preoperative biopsy, 25% of cases in both development and validation sets had a high-grade tumor. The ultrasound examiner assumed that there was cervical invasion in 17% (21/125) of cases and deep myometrial invasion in 41% (51/125).

Table S1 shows the univariate analysis of the 2D and 3D gray-scale and color Doppler ultrasound variables for the women with Grades 1 and 2 endometrial cancer according to preoperative biopsy included in the development set ($n = 94$). The retained variables for the objective model were preoperative grade and minimal tumor-free myometrium. The subjective model used subjective assessment of myometrial invasion, preoperative grade and age as variables.

The probability of having high-risk cancer is equal to $y1/(1 + e^{-z})$, where z is $-0.4468 + 1.2921 \times$ preoperative grading $-0.2292 \times$ minimal-free myometrium (for the objective model) and $-2.6276 + 1.1458 \times$ preoperative grading $+ 2.2514 \times$ subjective evaluation of myometrial invasion (for the subjective model).

Table 3 shows the results for predicting high-risk cancer for: a one-step strategy using Grade 3 only; the combined two-step strategies as described above; and the performance of the new ultrasound-based mathematical models and subjective assessment alone (when tested only on the preoperative Grade 1 and 2 patients). Of the 31 women classified as high risk according to preoperative grading, one woman was later found to be low risk

Table 1 Demographic and baseline characteristics of patients included in the total development set ($n = 125 = 94 + 31$) and the total validation set ($n = 211 = 163 + 48$) stratified by 'low-risk' and 'high-risk' cancer according to final pathology

Characteristic	Development set				Validation set			
	Total ($n = 125$)	Low risk ($n = 55$)	High risk ($n = 70$)	P	Total ($n = 211$)	Low risk ($n = 90$)	High risk ($n = 121$)	P
Age (years)	66 ± 10 66 (37–92)	63 ± 9 64 (37–82)	68 ± 10 67 (44–92)	0.006	65 ± 9 64 (41–89)	62 ± 9 61 (41–81)	66 ± 9 66 (44–89)	< 0.001
BMI*	29 ± 7 27 (19–55)	29 ± 6 27 (19–46)	29 ± 7 27 (20–55)	0.89	31 ± 8 30 (18–61)	32 ± 8 30 (18–55)	31 ± 7 30 (20–61)	0.34
Postmenopausal								
No	8 (6)	3 (5)	5 (7)	1	17 (8)	12 (13)	5 (4)	0.02
Yes	117 (94)	52 (95)	65 (93)		194 (92)	78 (87)	116 (96)	
Hormone treatment								
No	112 (90)	48 (87)	64 (91)	0.56	199 (94)	83 (92)	116 (96)	0.37
Yes	13 (10)	7 (13)	6 (9)		12 (6)	7 (8)	5 (4)	
Current low-potency estrogen use								
No	112 (90)	47 (85)	65 (93)	0.24	210 (99.5)	89 (99)	121 (100)	—
Yes	13 (10)	8 (15)	5 (7)		1 (0.5)	1 (1)	—	
Current tamoxifen use								
No	122 (98)	52 (95)	70 (100)	0.08	209 (99)	90 (100)	119 (98)	—
Yes	3 (2)	3 (5)	—		2 (1)	—	2 (2)	
Family history								
Breast cancer								
No	106 (85)	47 (85)	59 (84)	1	200 (95)	86 (96)	114 (94)	0.76
Yes	19 (15)	8 (15)	11 (16)		11 (5)	4 (4)	7 (6)	
Gynecological cancer								
No	107 (86)	45 (82)	62 (89)	0.31	198 (94)	80 (89)	118 (98)	0.02
Yes	18 (14)	10 (18)	8 (11)		13 (6)	10 (11)	3 (2)	
Diagnosis by:								
Pipelle biopsy	21 (19)	12 (27)	9 (14)	0.14	—	—	—	0.58
Hysteroscopy only	19 (17)	4 (9)	15 (23)		93 (46)	42 (48)	51 (44)	
D&C only	52 (48)	22 (49)	30 (47)		82 (40)	36 (41)	46 (40)	
Hysteroscopy and D&C	17 (16)	7 (16)	10 (16)		29 (14)	10 (11)	19 (16)	
No data/not performed/ other method (tru-cut biopsy/biopsy of meta)	16	10	6		7	2	5	

Values are given as n , n (%), mean ± SD or median (range). *Data for BMI missing in three low-risk and 12 high-risk cases in the development set. BMI, body mass index; D&C, dilatation and curettage.

according to final histology. Subjective assessment alone performed as well as any model. A two-step strategy with the 'subjective' or 'objective' model as the second step achieved a significantly higher sensitivity in comparison with preoperative grading only. All two-step strategies using models had similar specificities, indicating that ultrasound in the combination with grade will improve the correct identification of high-risk cases. The two-step strategies detected up to twice as many high-risk cases as did the one-step strategy based on grade only. Differences when using different cut-offs are explained below.

In Tables S2 and S3 we compared the P -values for the difference in sensitivity and specificity between the two-step strategy using models or subjective assessment, and between the use of different cut-offs in the development set. The only difference between the models and subjective assessment using the 0.5 cut-off was a significantly lower specificity for the objective model ($P = 0.03$), and when using the 0.4 cut-off both objective and subjective models had a lower specificity (both $P < 0.001$), the subjective model having a higher sensitivity ($P = 0.03$) compared with subjective assessment. Thus, the 0.5 cut-off achieved

a result more similar to subjective assessment, but the use of a 0.4 cut-off detected more high-risk cases at a cost of lower specificity.

Table 4 shows the external validation of preoperative grading only and the two-step strategies with our 'new' models, subjective assessment or previously published models (of Karlsson and coworkers and De Smet and coworkers) as the second step. As in Table 3 the performance is presented for the models individually or for the whole two-step strategy. Of the 48 women classified as high risk according to preoperative grading, four were later found to be low risk according to final histology. The ultrasound examiner assumed that there was cervical invasion in 11% (24/211) cases and deep myometrial invasion in 52% (109/211). The performance of the models in the validation set was very similar to the results from the development set.

Tables S4 and S5 compare the P -values for the differences in sensitivity and specificity between two-step strategies using new and old models as the second step and subjective assessment using different cut-off levels and subjective assessment for the validation set. The

Table 2 Histological characteristics of patients included in the total development set ($n = 125 = 94 + 31$) and the validation set ($n = 211 = 163 + 48$)

Characteristic	Development set			Validation set		
	Total ($n = 125$)	Grade 3 ($n = 40$)	Grade 1 or 2 ($n = 85$)	Total ($n = 211$)	Grade 3 ($n = 51$)	Grade 1 or 2 ($n = 160$)
Stage, FIGO 2009 ²²						
IA	64 (51)	11 (27.5)	53 (62)	106 (50)	17 (33)	89 (56)
IB	28 (22)	9 (22.5)	19 (22)	58 (27)	10 (20)	48 (30)
II	11 (9)	4 (10.0)	7 (8)	19 (9)	9 (18)	10 (6)
IIIA	3 (2)	3 (7.5)	—	6 (3)	3 (6)	3 (2)
IIIB	2 (2)	1 (2.5)	1 (1)	1 (<1)	1 (2)	—
IIIC	11 (9)	6 (15.0)	5 (6)	16 (8)	7 (14)	9 (6)
IVA	—	—	—	—	—	—
IVB	6 (5)	6 (15.0)	—	5 (2)	4 (8)	1 (<1)
		Development set ($n = 125$)		Validation set ($n = 211$)		
Type and grade preoperatively						
Endometrioid						
Grade 1		65 (52)			89 (42)	
Grade 2		29 (23)			74 (35)	
Grade 3		20 (16)			31 (15)	
Non-endometrioid		11 (9)			17 (8)	
Type and grade postoperatively						
Endometrioid						
Grade 1		51 (41)			100 (47)	
Grade 2		34 (27)			60 (28)	
Grade 3		19 (15)			35 (16)	
Non-endometrioid		21 (17)*			16 (8)†	

Values are given as n (%). *Non-endometrioid histotypes, after final surgery, in the development set (four carcinosarcoma, three clear cell, three mixed, three seropapillary and eight serous). †Non-endometrioid histotypes, after final surgery, in the validation set (four clear cell, two mixed, seven serous, two undifferentiated, one squamous).

sensitivity for any of the new models and for subjective assessment performed significantly better compared with the old models at the same level of specificity.

In Figure 3a and b, pie charts present and compare the performance of all strategies on the development and the validation set.

Figure 4 shows the percentage of additional women with a correct diagnosis compared with the one-step strategy using preoperative Grade 3 only in the validation set. The subjective models and subjective assessment improved the accuracy by 16%. In the development set the accuracy was improved by up to 29% (for subjective assessment and the subjective model (at the cut-off of 0.5)).

DISCUSSION

Good preoperative risk stratification is of utmost importance. Surgical staging to identify high-risk cancers and to decide whether to proceed with a lymphadenectomy is based on several factors: histological tumor type; grade of tumor differentiation; depth of myometrial invasion; and cervical extension. The first two factors can be determined on the preoperative endometrial biopsy, although the tumor grade might have been underestimated as becomes clear on the final microscopic analysis of the hysterectomy specimen. However, depth of myometrial invasion and cervical extension to a certain extent can only be assessed during surgery and therefore are not available to guide appropriate referral²³.

In this study we prospectively developed and tested new mathematical models aiming to preoperatively identify women with high-risk endometrial cancer. We found that a two-step strategy combining Grade 3 with either one of the models as a second step or subjective assessment detected up to almost twice as many high-risk cases as preoperative grade alone (78–83% for two-step strategies *vs* 36% for preoperative grade alone) at the comparable level of specificity (68% for the objective model, 72% for the subjective model and 71% for subjective assessment) *vs* 96% (validation set, cut-off level = 0.50 for both models). Subjective assessment performed as well as any of the models. Based on the preoperative grading, 75% or 77% (development and validation sets, respectively) of the women were considered as low risk. Among these low-risk patients, 42% (40/94) and 47% (77/163) were classified as high risk according to the final surgical staging, indicating a need for repeat surgery with lymphadenectomy or complementary treatment with radiochemotherapy. By introducing the two-step strategy one could improve the correct classification of the patients by up to 16% (Figure 4). Looking at the whole group, preoperative grading only would miss 64% of the high-risk patients, the models would miss 17–22% of the high-risk patients and subjective assessment would miss 17%.

We tested two different cut-off levels for the models to determine which would yield the most favorable clinical outcome in terms of trade-off between sensitivity and specificity. The 0.5 cut-off gave results that were very

Table 3 Ability of the two-step strategy to detect high-risk endometrial cancer through combining preoperative Grade 3 with ultrasound-based mathematical models or subjective assessment in the development set ($n = 125$)

Preoperative Grade 3 ($n = 31$)	Model	Preoperative Grade 1 or 2 ($n = 94$)	Total ($n = 125$)
PPV = 30/31	Objective model		
	AUC	0.76 (0.66–0.86)	
	Cut-off 0.5		
	Sensitivity (%)	63 (46–77) [25/40]	79 (68–87) [55/70]
	Specificity (%)	74 (60–85) [40/54]	73 (60–83) [40/55]
	LR+	2.41 (1.45–4.02)	2.88 (1.84–4.51)
	LR–	0.51 (0.33–0.78)	0.30 (0.18–0.48)
	Cut-off 0.4		
	Sensitivity (%)	70 (53–83) [28/40]	83 (72–90) [58/70]
	Specificity (%)	61 (47–74) [33/54]	60 (47–72) [33/55]
	LR+	1.80 (1.22–2.66)	2.07 (1.47–2.91)
	LR–	0.49 (0.29–0.83)	0.29 (0.16–0.50)
	Subjective model		
	AUC	0.78 (0.68–0.87)	
	Cut-off 0.5		
	Sensitivity (%)	63 (46–77) [25/40]	79 (68–87) [55/70]
	Specificity (%)	85 (73–83) [46/54]	84 (72–91) [46/55]
	LR+	4.22 (2.13–8.35)	4.80 (2.61–8.84)
	LR–	0.44 (0.29–0.67)	0.26 (0.16–0.41)
	Cut-off 0.4		
	Sensitivity (%)	75 (59–87) [30/40]	86 (76–92) [60/70]
	Specificity (%)	65 (51–77) [35/54]	64 (50–75) [35/55]
	LR+	2.13 (1.42–3.19)	2.36 (1.64–3.39)
	LR–	0.39 (0.22–0.68)	0.22 (0.12–0.41)
	Subjective impression		
	Sensitivity (%)	63 (46–77) [25/40]	79 (68–87) [55/70]
	Specificity (%)	85 (73–93) [46/54]	84 (72–91) [46/55]
LR+	4.22 (2.13–8.35)	4.80 (2.61–8.84)	
LR–	0.44 (0.29–0.67)	0.26 (0.16–0.41)	
All low risk			
Sensitivity (%)	0 (0–9) [0/40]	43 (32–55) [30/70]	
Specificity (%)	100 (93–100) [54/54]	98 (90–100) [54/55]	
LR+		23.57 (3.32–167.48)	
LR–		0.58 (0.47–0.72)	

Values in round parentheses are 95% CI and those in square parentheses are n/n . AUC, area under the curve; LR+, positive likelihood ratio; LR–, negative likelihood ratio; PPV, positive predictive value.

close to that of subjective impression in both development and validation sets. There seems to be no benefit in use of the 0.4 cut-off.

The variables selected by the objective model were preoperative grade (Grade 1 or 2) and minimal tumor-free myometrium, which are not directly related to the subjective assessment of the examiner.

In the past, several attempts have been made to improve the current preoperative risk stratification. The aim of Karlsson's criteria was to predict deep myometrial infiltration, not high-risk cancer patients. Using his criteria, as described above, he achieved a sensitivity of 79% and a specificity of 100%¹⁷. Because Karlsson also described cervical invasion as one of the criteria, we could assess Karlsson's criteria as a two-step strategy also for the prediction of high-risk cancers. Alcazar *et al.* assessed myometrial infiltration by assessing stored 3D volumes. In each plane the minimal tumor-free myometrial thickness was measured. The study showed that the best cut-off to predict more than 50% myometrial infiltration was a minimal tumor-free myometrial margin of 9 mm, resulting

in a sensitivity of 100% and a specificity of 61%²⁴. De Smet *et al.* developed a logistic regression model with degree of differentiation, number of fibroids, endometrial thickness and tumor volume as variables. He achieved, after validation, a sensitivity of 50% to predict deep infiltration and a specificity of 75%¹⁸.

We compared our new models with those previously developed models (Karlsson and De Smet) and found that the new models performed significantly better, the sensitivity being significantly better for the new model at the same level of specificity. Nevertheless, there is still room for improvement of the performance of our models because we did not achieve optimal results for sensitivity and specificity. Perhaps new biochemical tumor markers, new histological tumor markers or new ultrasound techniques can improve this performance, but we conclude that, at present, the current 3D ultrasound variables are not able to improve performance and 2D ultrasound assessment is sufficient at this stage.

Because De Smet *et al.* did not assess cervical invasion, which inevitably upgrades the stage to Stage II, the

Table 4 External validation of the new and previously developed models and subjective assessment (validation set, $n = 211$)

<i>Preoperative Grade 3 (n = 48)</i>	<i>Model</i>	<i>Preoperative Grade 1 or 2 (n = 163)</i>	<i>Total (n = 211)</i>
PPV = 44/48	Objective model		
	AUC	0.73 (0.66–0.81)	
	Cut-off 0.5		
	Sensitivity (%)	65 (53–75) [50/77]	78 (69–84) [94/121]
	Specificity (%)	71 (60–80) [61/86]	68 (58–77) [61/90]
	LR+	2.23 (1.55–3.23)	2.41 (1.76–3.30)
	LR–	0.49 (0.35–0.69)	0.33 (0.23–0.47)
	Cut-off 0.40		
	Sensitivity (%)	73 (61–82) [56/77]	83 (75–88) [100/121]
	Specificity (%)	58 (47–69) [50/86]	56 (45–65) [50/90]
	LR+	1.74 (1.31–2.31)	1.86 (1.46–2.38)
	LR–	0.47 (0.31–0.70)	0.31 (0.20–0.48)
	Subjective model		
	AUC	0.77 (0.70–0.85)	
	Cut-off 0.5		
	Sensitivity (%)	71 (60–81) [55/77]	82 (74–88) [99/121]
	Specificity (%)	76 (65–84) [65/86]	72 (62–80) [65/90]
	LR+	2.93 (1.97–4.35)	2.95 (2.09–4.15)
	LR–	0.38 (0.26–0.55)	0.25 (0.17–0.38)
	Cut-off 0.4		
	Sensitivity (%)	71 (60–81) [55/77]	82 (74–88) [99/121]
	Specificity (%)	76 (65–84) [65/86]	72 (62–80) [65/90]
	LR+	2.93 (1.97–4.35)	2.95 (2.09–4.15)
	LR–	0.38 (0.26–0.55)	0.25 (0.17–0.38)
	Subjective assessment		
	Sensitivity (%)	73 (62–81) [56/77]	83 (75–88) [100/121]
	Specificity (%)	74 (64–83) [64/86]	71 (61–79) [64/90]
	LR+	2.84 (1.93–4.18)	2.86 (2.05–4.00)
	LR–	0.37 (0.25–0.54)	0.24 (0.16–0.37)
	De Smet		
	Sensitivity (%)	51 (40–62) [39/77]	69 (60–76) [83/121]
	Specificity (%)	79 (69–87) [68/86]	76 (66–83) [68/90]
	LR+	2.42 (1.52–3.86)	2.81 (1.91–4.12)
	LR–	0.62 (0.49–0.80)	0.42 (0.31–0.56)
	Karlsson		
	Sensitivity (%)	47 (42–65) [36/77]	66 (57–74) [80/121]
	Specificity (%)	78 (68–86) [67/86]	74 (65–82) [67/90]
	LR+	2.12 (1.33–3.36)	2.59 (1.78–3.76)
	LR–	0.68 (0.54–0.87)	0.46 (0.35–0.60)
	All low risk		
	Sensitivity (%)	0 (0–5) [0/77]	36 (28–45) [44/121]
	Specificity (%)	100 (96–100) [86/86]	96 (89–98) [86/90]
	LR+		8.18 (3.05–21.94)
	LR–		0.67 (0.58–0.77)

Values in round parentheses are 95% CI and those in square parentheses are n/n . AUC, area under the curve; LR+, positive likelihood ratio; LR–, negative likelihood ratio; PPV, positive predictive value.

comparison between the two-step strategy using De Smet's model as the second step and the other two-step strategies is not completely fair and the two-step strategy with De Smet's model might be performing less well, for this reason. Another weakness in this study is that a lymphadenectomy was not performed in all patients because patients were treated according to local protocols and, in several centers, patients with low-risk cancer do not undergo a lymphadenectomy. Nevertheless, the new FIGO staging requires lymphadenectomy and therefore the staging of the patients might be underestimated.

In several centers MRI is the standard preoperative imaging for risk stratification, but this is certainly not the

case in all countries; therefore, we decided to develop and evaluate strategies with ultrasound-based models rather than with MRI, although a comparison with MRI would have been interesting.

It remains to be shown whether a two-step strategy containing subjective ultrasound assessment or objective ultrasound measurements will have the highest reproducibility and if any of the strategies work as well in the hands of less-experienced examiners.

The issues that need to be considered when we compare the debate of the preoperative prediction of malignancy in adnexal masses with that of the prediction of high-risk endometrial cancers, are completely different.

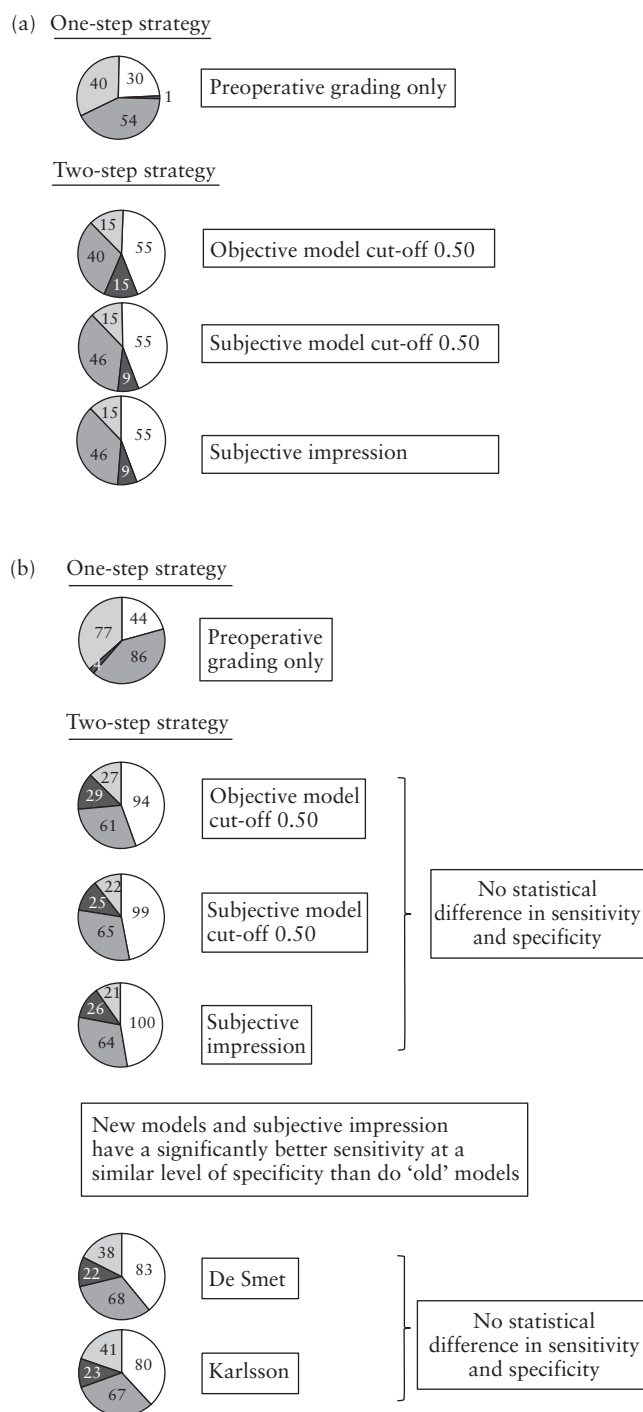


Figure 3 Pie charts demonstrating performance of one-step and two-step strategies on the development set ($n = 125$) (a) and on the whole validation set ($n = 211$) (b). □, true positive; ■, false positive; ▤, true negative; ◻, false negative.

The main reason is that there is no general international consensus on the treatment strategy of patients with endometrial cancers, more specifically on which patients should be treated with a lymphadenectomy and which patients should receive adjuvant radiotherapy. Therefore, depending on the clinical setting of the local hospital and the local protocol that is followed, the cut-off that is used by the model could be changed. Hospitals that offer easy access to a team with special skills in lymphadenectomy

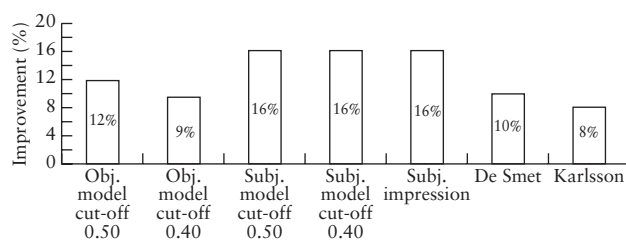


Figure 4 Percentage improvement, using validation set ($n = 211$), in preoperative classification of endometrial cancers using ultrasound-based models (objective (Obj.) and subjective (Subj.) models with two different cut-offs, subjective impression, and the older models of De Smet and Karlsson), compared with preoperative grading only.

could opt for the cut-off that gives the highest sensitivity for an acceptable false-positive rate, thus reducing the number of patients who undergo unnecessary radiotherapy. On the other hand, patients with endometrial cancer are quite often obese with a substantial degree of comorbidity, making lymphadenectomy a technically difficult procedure with substantial risks of complications.

In conclusion, based on this study we have shown that a two-step strategy combining preoperative grade and ultrasound-based mathematical models can help us identify twice as many women with high-risk cancer compared with a strategy based on preoperative grade alone, and can improve the correct classification of patients by 29%. Further research is needed to evaluate whether referral based on this adapted strategy will improve patient outcome.

ACKNOWLEDGMENT

The external validation study was supported by the Internal Grant Agency of the Ministry of Health, the Czech Republic (grant NT 13070), and by Charles University in Prague (UNCE 204024 and PRVOUK-P27/LF1/1).

REFERENCES

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin* 2011; **61**: 69–90.
2. Lewin S, Herzog T, Barrena Medel N, Deutsch I, Burke W, Sun X, Wright J. Comparative performance of the 2009 international Federation of gynecology and obstetrics' staging system for uterine corpus cancer. *Obstet Gynecol* 2010; **116**: 1141–1149.
3. Morrow CP, Bundy BN, Kurman RJ, Creasman WT, Heller P, Homesley HD, Graham JE Relationship between surgical-pathological risk factors and outcome in clinical stage I and II carcinoma of the endometrium: a Gynecologic Oncology Group study. *Gynecol Oncol* 1991; **40**: 55–65.
4. Brown AK, Madom L, Moore R, Granai CO, DiSilvestro P. The prognostic significance of lower uterine segment involvement in surgically staged endometrial cancer patients with negative nodes. *Gynecol Oncol* 2007; **105**: 55–58.
5. Nag S, Erickson B, Parikh S, Gupta N, Varia M, Glasgow G. The American Brachytherapy Society recommendations for high-dose-rate brachytherapy for carcinoma of the endometrium. *Int J Radiat Oncol Biol Phys* 2000; **48**: 779.

6. Kitchener H, Swart AM, Qian Q, Amos C, Parmar MK. Efficacy of systematic pelvic lymphadenectomy in endometrial cancer (MRC ASTEC trial): a randomised study. *Lancet* 2009; **373**: 125–136.
7. Blake P, Swart AM, Orton J, Kitchener H, Whelan T, Lukka H, Eisenhauer E, Bacon M, Tu D, Parmar MK, Amos C, Murray C, Qian W. Adjuvant external beam radiotherapy in the treatment of endometrial cancer (MRC ASTEC and NCIC CTG EN.5 randomised trials): pooled trial results, systematic review, and meta-analysis. *Lancet* 2009; **373**: 137–146.
8. Akbayir O, Corbacioglu A, Numanoglu C, Guleroglu FY, Ulker V, Akyol A, Guraslan B, Odabasi E. Preoperative assessment of myometrial and cervical invasion in endometrial carcinoma by transvaginal ultrasound. *Gynecol Oncol* 2011; **122**: 600–603.
9. Alcazar JL, Galvan R, Albela S, Martinez S, Pahisa J, Jurado M, Lopez-Garcia G. Assessing myometrial infiltration by endometrial cancer: uterine virtual navigation with three-dimensional US. *Radiology* 2009; **250**: 776–783.
10. Savelli L, Ceccarini M, Ludovisi M, Fruscella E, De Iaco PA, Salizzoni E, Mabrouk M, Manfredi R, Testa AC, Ferrandina G. Preoperative local staging of endometrial cancer: transvaginal sonography vs. magnetic resonance imaging. *Ultrasound Obstet Gynecol* 2008; **31**: 560–566.
11. Akbayir O, Corbacioglu A, Goksedef BP, Numanoglu C, Akca A, Guraslan H, Bakir LV, Cetin A. The novel criteria for predicting pelvic lymph node metastasis in endometrioid adenocarcinoma of endometrium. *Gynecol Oncol* 2012; **125**: 400–403.
12. Sawicki W, Spiewankiewicz B, Stelmachów J, Cendrowski K. The value of ultrasonography in preoperative assessment of selected prognostic factors in endometrial cancer. *Eur J Gynaecol Oncol* 2003; **24**: 293–298.
13. Pete I, Godény M, Tóth E, Radó J, Pete B, Pulay T. Prediction of cervical infiltration in Stage II endometrial cancer by different preoperative evaluation techniques (D&C, US, CT, MRI). *Eur J Gynaecol Oncol* 2003; **24**: 517–522.
14. Haldorsen IS, Berg A, Werner HM, Magnussen IJ, Helland H, Salvesen OO, Trovik J, Salvesen HB. Magnetic resonance imaging performs better than endocervical curettage for preoperative prediction of cervical stromal invasion in endometrial carcinomas. *Gynecol Oncol* 2012; **126**: 413–418.
15. Savelli L, Ceccarini M, Ludovisi M, Fruscella E, De Iaco PA, Salizzoni E, Mabrouk M, Manfredi R, Testa AC, Ferrandina G. Preoperative local staging of endometrial cancer: transvaginal sonography vs. magnetic resonance imaging. *Ultrasound Obstet Gynecol* 2008; **31**: 560–566.
16. Sala E, Rockall AG, Freeman SJ, Mitchell DG, Reinhold C. The added role of MR imaging in treatment stratification of patients with gynecologic malignancies: what the radiologist needs to know. *Radiology* 2013; **266**: 717–740.
17. Karlsson B, Norström A, Granberg S, Wikland M. The use of endovaginal ultrasound to diagnose invasion of endometrial cancer. *Ultrasound Obstet Gynecol* 1992; **2**: 35–39.
18. De Smet F, De Brabanter J, Van den Bosch T, Pochet N, Amant F, Van Holsbeke C, Moerman P, De Moor B, Vergote I, Timmerman D. New models to predict depth of infiltration in endometrial carcinoma based on transvaginal sonography. *Ultrasound Obstet Gynecol* 2006; **27**: 664–671.
19. Leone F, Timmerman D, Bourne T, Valentin L, Epstein E, Goldstein S, Marret H, Parsons A, Gull B, Istre O, Sepulveda W, Ferrazzi E, Van den Bosch T. Terms, definitions and measurements to describe the sonographic features of the endometrium and intrauterine lesions: a consensus opinion from the International Endometrial Tumor Analysis (IETA) group. *Ultrasound Obstet Gynecol* 2010; **35**: 103–112.
20. Timmerman D, Valentin L, Bourne TH, Collins WP, Verrelst H, Vergote I. Terms, definitions and measurements to describe the sonographic features of adnexal tumors: a consensus opinion from the International Ovarian Tumor Analysis (IOTA) Group. *Ultrasound Obstet Gynecol* 2000; **16**: 500–505.
21. World Health Organization Classification of Tumours. Tumours of the uterine corpus. In *Pathology and Genetics. Tumours of the Breast and Female Genital Organs*, Tavassoli FA, Devilee P (eds). IARC Press: Lyon, 2003; 217–287.
22. Pecorelli S. Revised FIGO staging for carcinoma of the vulva, cervix, and endometrium. *Int J Gynaecol Obstet* 2009; **105**: 103–104.
23. Ramondetta L, Burke T, Broaddus R, Jhingran A. Treatment of endometrial cancer. In *The MD Anderson Cancer Care Series: Gynecologic Cancer*, Eifel PJ, Gershenson DM, Kavanagh JJ, Silva EG (eds). Springer-Verlag: New York, NY, 2006; 152–153.
24. Alcazar JL, Galvan R, Albela S, Martinez S, Pahisa J, Jurado M, Lopez-Garcia G. Assessing myometrial infiltration by endometrial cancer: uterine virtual navigation with three-dimensional US. *Radiology* 2009; **250**: 776–783.

SUPPORTING INFORMATION ON THE INTERNET

The following supporting information may be found in the online version of this article:



Table S1 Univariate analysis of the two-dimensional (2D) and three-dimensional (3D) gray-scale and color Doppler ultrasound variables of the endometrial cancers included in the model building dataset ($n = 94$)

Tables S2 and S3 Difference in sensitivity (Table S2) and specificity (Table S3) between two-step strategies using models as the second step and subjective assessment, and according to different cut-offs (values are given as P) (development set, $n = 125$)

Tables S4 and S5 Difference in sensitivity (Table S4) and specificity (Table S5) between two-step strategies using models as the second step and subjective assessment, and according to different cut-offs (values are given as P) (validation set, $n = 211$)