The use of ultrasound in the prediction of endometrial cancer in women with postmenopausal bleeding

Opolskiene, Gina

2010

Department of Clinical Sciences, Malmö
Obstetric, Gynaecological and Prenatal Ultrasound Research Unit
Lund University, Sweden

The Use of Ultrasound in the Prediction of Endometrial Cancer in Women with Postmenopausal Bleeding

Gina Opolskiene

Academic Dissertation

With permission of the Faculty of Medicine of Lund University,
to be presented for public examination at the Department
of Obstetrics and Gynaecology, Skåne University Hospital, Malmö

November 26, 2010 at 9 a.m.

Faculty opponent
Professor Christoff Brezinka
Department of Obstetrics and Gynecology, Innsbruck University Hospital,
Anichstrasse 35, A-6020, Innsbruck, Austria

HÄR SKA SPIKBLADET IN.

To my daughters Vita and Jorile

"I will prescribe regimens for the good of my patients according to my ability and my judgment and never do harm to anyone…"
(from the Hippocratic Oath**)**

Preface

As a medical student, I was taught by more senior gynaecologists that the first measure in cases of postmenopausal bleeding was dilatation and curettage, as postmenopausal bleeding could be a sign of endometrial cancer. However, the cause of postmenopausal bleeding is often benign.

Postmenopausal women are by definition a special group of women in that they are normally over 50 years of age. The older women in this group may also suffer from comorbidity. Therefore, any interventional procedure is associated with a risk. With the introduction of ultrasound into gynaecology it became possible to separate women with postmenopausal bleeding into a high- and low-risk group for endometrial cancer depending on the endometrial thickness.

The aim of the work presented in this thesis is to identify the ultrasound methods and ultrasound features that are most useful for the prediction of endometrial cancer with the purpose of individualizing treatment and avoiding hazardous interventions for patients with postmenopausal bleeding and endometrial thickness ≥ 4.5 mm.

List of papers

This thesis is based on the following papers, which will be referred to in the text by their roman numerals. The papers are appended at the end of the thesis. Permission for reprinting is granted by publishers.

I. Opolskiene G, Sladkevicius P & Valentin L
    Ultrasound assessment of endometrial morphology and vascularity to predict endometrial malignancy in women with postmenopausal bleeding and sonographic endometrial thickness ≥ 4.5 mm
    Ultrasound Obstet Gynecol 2007; 30: 332-340

II. Opolskiene G, Sladkevicius P & Valentin L
    Prediction of endometrial malignancy in women with postmenopausal bleeding and sonographic endometrial thickness ≥ 4.5 mm
    Submitted

III. Opolskiene G, Sladkevicius P, Jokubkiene L & Valentin L
    Three-dimensional ultrasound imaging for discrimination between benign and malignant endometrium in women with postmenopausal bleeding and sonographic endometrial thickness of at least 4.5 mm
    Ultrasound Obstet Gynecol 2010; 35: 94-102

IV. Opolskiene G, Sladkevicius P & Valentin L
    Two- and three-dimensional saline contrast sonohysterography: interobserver agreement, agreement with hysteroscopy and diagnosis of endometrial malignancy
    Ultrasound Obstet Gynecol 2009; 33: 574-582

Abbreviations and definitions

AUC	area under the curve
CI	confidence interval
D&C	dilatation and curettage
FI	flow index
HRT	hormonal replacement therapy
LR	likelihood ratio
LR+	positive likelihood ratio
LR-	negative likelihood ratio
ROC	receiver operating characteristic
SCSH	saline contrast sonohysterography
VAS	visual analogue scale
VI	vascularization index
VFI	vascularization-flow index
2D	two-dimensional
3D	three-dimensional

Contents

Background

Endometrial cancer

Endometrial cancer is one of the most common forms of gynaecological cancer in the developed countries, and the incidence is rising (Amant et al., 2005). Endometrial cancer usually affects postmenopausal women, and is rare before the age of 40. Only less than 20% of endometrial cancers occur before menopause (Engelsen et al., 2009).

The prognosis is generally good, as the cancer is often diagnosed while the tumour is confined to the uterine corpus. The overall survival rate is about 75% (Danforth's Obstetrics and Gynecology 2003), and depends on the stage and subtype of the cancer. Endometrial cancer is divided in two subtypes: type I, which accounts for about 80% of all endometrial cancers, and is associated with endometrioid histology and a good prognosis, and type II, associated with non-endometrioid histology (clear-cell, serous papillary or mucinous carcinoma) and a poor prognosis. However, early-stage endometrial cancer and type I endometrial cancer recur in up to 20% of cases (Engelsen et al., 2009).

The risk of endometrial cancer increases with obesity, diabetes, hypertension, nulliparity, hormonal replacement therapy (HRT) and tamoxifen therapy. Increasing age and a family history of hereditary non-polyposis colorectal cancer syndrome are also risk factors for endometrial cancer (Amant et al., 2005). Endometrial hyperplasia, especially atypical, is a precursor of type I endometrial cancer (Amant et al., 2005).

Postmenopausal bleeding

There are many benign causes of postmenopausal bleeding, including atrophic endo-metrium (50%), endometrial hyperplasia (13%) and endometrial polyps (10%), among others (Ferrazzi et al, 1996). However, there is also about a 1% probability of cervical cancer (Karlsson et al., 1995) and about a 10% probability of endometrial cancer in women with postmenopausal bleeding (Smith–Bindman et al., 1989, Ferrazzi et al., 1996). In most cases, the first sign of endometrial cancer is postmenopausal bleeding.

Endometrial thickness

The normal postmenopausal endometrium is thin, uniform and consists of the basalis layer, which is less than 1 mm thick anatomically (Parsons, 1998). Sonographic transvaginal measurements of the endometrium have a slight tendency to overestimate the anatomical endometrial thickness (Saha et al., 2004), but normal unstimulated postmenopausal endometrium does not exceed 4-5 mm in thickness when measured with ultrasound (Parsons, 1998, Karlsson et al., 1995). A thin, 1-2 mm hypoechoic rim (also called the subendometrium) between the echogenic endometrium and the myometrium, a halo, usually surrounds the normal postmenopausal endometrium (Sheth et al., 1993). This halo must not be included in the measurement of endometrial thickness (Epstein & Valentin, 2004), as it represents a network of capillaries and veins (Fleischer et al., 1986).

The sonographic measurement of endometrial thickness makes it possible to divide women with postmenopausal bleeding into a low-risk group and a high-risk group: if the sonographic endometrial thickness is less than 4.5 mm the risk of endometrial malignancy is low, if it is 4.5 mm or more, the risk of endometrial malignancy is high (Smith-Bindman et al, 1998). In the low-risk group, expectant management is possible, no endometrial sample is taken, but women are told to return if bleeding recurs (Gull et al., 2000, Goldstein et al., 2001, Gupta et al., 2002). Women with postmenopausal bleeding and a thick endometrium must undergo endometrial sampling because of the high risk of endometrial cancer.

Endometrial sampling

Endometrial sampling is used to detect endometrial malignancy in cases of postmenopausal bleeding. Outpatient endometrial sampling using a simple device, for example, a Pipelle de Cornier® (Prodimed, Neuilly-en-Thelle, France) or Endorette® (Medscand AB, Malmö, Sweden) is a simple procedure. However, the failure rate of outpatient endometrial sampling after menopause is high (16-18%), most often because of cervical stenosis (Epstein et al., 2001, De Silva et al., 1997). Moreover, endometrial samples obtained with an outpatient sampling device often contain insufficient material, and endometrial cancer may be missed (Goldstein, 2009, Smith–Bindman et al., 1998). If outpatient sampling fails, the woman must undergo a dilatation and curettage (D&C) or

hysteroscopy with analgesia or under anaesthesia. However, a substantial proportion of postmenopausal women are at high operative risk, and these procedures should only be carried out when absolutely necessary. In problematic cases, calculating the individual risk of endometrial malignancy using logistic regression models based on ultrasound variables could help to tailor the management of women with postmenopausal bleeding and thick endometrium.

Ultrasound methods for the assessment of thick postmenopausal endometrium
The endometrium is defined as being thick when it is ≥ 4.5 mm thick, when measured sonographically. Various ultrasound methods can be used to predict endometrial cancer and to estimate the individual risk of malignancy in women with postmenopausal bleeding and thick endometrium.

Grey-scale ultrasound for discrimination between benign and malignant endometrium
The sonographic measurement of endometrial thickness is very important in the group of women with a thick (≥ 4.5 mm) endometrium and postmenopausal bleeding, as the risk of endometrial cancer increases with every additional millimetre (Epstein et al., 2002). In addition to measuring the endometrial thickness, the morphology can be evaluated using two-dimensional (2D) grey-scale ultrasound. Suspicious sonomorphological signs of endometrial cancer are:
- heterogeneous endometrial echogenicity (Epstein & Valentin, 2006) and
- an irregular endometrial–myometrial border (Randelzhofer et al., 2002).

The echogenicity of the endometrium is related to the amount of mucin in the endometrial glands (Fleischer et al., 1986). Elevated amounts of mucin in the endometrium, as well as areas of haemorrhage and necrosis (Sheth et al., 1993) may explain the heterogeneous endometrial echogenicity and indicate endometrial cancer. An irregular endometrial–myometrial border (endometrial–myometrial junction, according to Leone et al., 2010) reflects the invasion of endometrial cancer into the myometrium.
Sonomorphological signs associated with benign endometrial pathology are:
- homogeneous endometrial echogenicity (Randelzhofer et al., 2002),

- regular endometrial–myometrial border (Randelzhofer et al., 2002),
- the presence of a middle echo (Weigel et al., 1995)
- the presence of hyperechoic lines separating the endometrial echo from the myometrium (Fong et al., 2003) and
- a cystic endometrial structure (Fong et al., 2003).

Hyperechoic lines (bright lines, according to Leone et al., 2010) and cystic endometrial structure are often seen in cases of endometrial polyps (Fong et al., 2003). The presence of a middle echo, i.e a line separating the two endometrial layers, signifies atrophic or non-pathological endometrium (Weigel, 1995), but the presence of a middle echo is a rather rare finding in postmenopausal endometrium (Randelzhofer, 2002).

Power Doppler ultrasound for discrimination between benign and malignant endometrium
Rubin and Adler introduced power Doppler ultrasound for radiological examinations in 1993 (Rubin & Adler, 1993). Power Doppler sonography is a technique that displays the strength of the Doppler signal in colour in the region of interest, and it reflects the number of red blood cells in the blood vessel (Martinoli et al., 1998). The specific features of this technique are that it is highly sensitive to blood flow and relatively independent of insonation angle (Martinoli et al., 1998). This results in good visualization of vessel morphology. Power Doppler ultrasound can be used to examine small, tortuous vessels and visualize their branching. The disadvantage of power Doppler ultrasound is its high sensitivity to artefacts such as tissue motion. This sometimes causes problems during examination and limits the application of the technique (Martinoli et al., 1998).

Increased angiogenesis, i.e. the formation of new capillaries, and increased microvascular density in the tissue are signs of a tumour (Fleischer et al., 2002, Abulafia & Sherer, 1999). The following vessel characteristics as assessed by power Doppler ultrasound are found more often in cases of endometrial cancer:
- irregular branching of vessels (Epstein & Valentin, 2006) and
- an increased number of vessels (Alcazar et al., 2003).

Alcazar et al. (2003) proposed three vascular patterns for the classification of sonographic endometrial vascularity:

A. a multiple vessel pattern, characteristic of endometrial cancer,

B. a single vessel pattern, characteristic of endometrial polyps, and

C. a scattered vascular pattern, characteristic of endometrial hyperplasia.

The degree of endometrial vascularity can be estimated subjectively or objectively by quantifying the colour content of the endometrial scan obtained using power Doppler ultrasound. This can be done subjectively using a visual analogue scale (VAS), usually graded from 0 to 100. Different computer programs are used for objective quantification. These programs capture the selected image, subtract the background grey scale pixels and summarize the number and values of the remaining colour pixels to quantify the colour content, which is related to blood flow (Martinoli et al., 1998). High colour content (abundant colour, according to Leone et al., 2010) indicates an increased risk of endometrial cancer (Epstein et al., 2002).

Three-dimensional (3D) power Doppler ultrasound for discrimination between benign and malignant endometrium

3D power Doppler ultrasound is a new technique that involves capturing a series of sequential images while the transducer is moved in a predictable manner (Martinoli et al., 1998). This technique allows a volume of the organ of interest to be acquired, which can be analysed off-line using appropriate software (often Virtual Organ Computer-Aided Analysis, VOCAL™). Using the VOCAL program it is possible to calculate the volume of the tissue being investigated and three power Doppler indices of vascularity: the vascularization index (VI), the flow index (FI) and the vascularization flow index (VFI) of the organ of interest. VI is the ratio of coloured voxels to all voxels, expressed as a percentage, and it reflects the density of vessels. FI is the sum of the weighted coloured voxels divided by the number of coloured voxels, and reflects the number of blood corpuscles flowing in the vessels. VFI is the sum of the weighted coloured voxels divided by the total number of voxels, and reflects both the density of vessels and the number of blood corpuscles (Pairleitner et al., 1999). Thus, the vascularity of the entire endometrium can be characterized.

It has been reported that high endometrial volume and high indices of vascularity are associated with endometrial cancer, and can thus be used to predict endometrial malignancy (Alcazar & Galvan, 2009, Odeh et al., 2007).

Saline contrast sonohysterography (SCSH) for discrimination between benign and malignant endometrium

SCSH, also called hydrosonography or saline infusion sonography, is a complement to standard transvaginal ultrasound examination. It is a simple, usually painless procedure, which requires a speculum, a plastic catheter, and a 20 ml syringe containing sterile saline, and it should be available at every clinic where pelvic ultrasound is performed (Parsons, 1998).

During SCSH a sterile plastic catheter (insemination catheters, baby feeding catheters or similar catheters) is inserted through the cervical canal, and the uterine cavity is distended by the injection of sterile saline. This makes it possible to evaluate the whole uterine cavity during the ultrasound examination. A sterile hydroxyethyl cellulose gel can be used instead of saline (Exalto et al., 2007). SCSH allows accurate discrimination between focal endometrial lesions and global endometrial thickening (Parsons, 1998), and can replace diagnostic hysteroscopy (Widrich et al., 1996).

Sonohysterographic features indicating endometrial cancer are:
- irregular thickening of the endometrium with variable echogenic texture, irregular endometrial–myometrial border (Parsons & Lense, 1993),
- difficulties with distending the uterine cavity during SCSH (Laifer-Narin et al., 1999, Bree et al., 2000, Epstein et al., 2001), and
- an irregular surface of a focal lesion (Bronz et al., 1997, Epstein & Valentin, 2006).

In addition to 2D SCSH, 3D SCSH can be performed, and some authors have suggested that 3D SCSH may be superior to 2D SCSH for the detection of intrauterine anomalies (La Torre et al, 1999, Bonilla-Musoles et al., 1997). One problem with SCSH is that the failure

rate after menopause is much higher than in premenopausal women (10.2-16.8% versus 5.0-13.5%), often because of cervical stenosis (De Kroon et al., 2003).

Concern has been expressed about performing SCSH when transvaginal ultrasound strongly suggests endometrial cancer (Amant et al., 2005, Dessole et al., 2006), as there is a small risk of malignant cells finding their way into the abdominal cavity via the fallopian tubes (Alcazar et al., 2000). Other studies found that SCSH does not increase the risk of dissemination of malignant cells (Takac et al., 2008) and that the use of SCSH is justified (Berre et al., 2008). Before 2009, the finding of malignant cells in the peritoneal cavity using the FIGO staging system (1988) classified endometrial cancer as stage IIIA (Shepherd, 1989). In 2009, the FIGO staging system was changed, such that the finding of malignant cells in peritoneal fluid is reported separately without changing the stage (Pecorelli, 2009). Obermair et al. (2000) reported similar disease-free, short-term survival in patients with endometrial cancer confined to the uterus irrespective of whether or not the women had undergone hysteroscopy before laparotomy. This was despite positive peritoneal cytology findings being more common (but not statistically significantly so) in patients who had undergone hysteroscopy.

Aims of the studies

The specific aims of the work described in this thesis are given below:

- To determine which grey-scale and power Doppler ultrasound variables are useful in discriminating between benign and malignant endometrium in women with postmenopausal bleeding and a sonographic endometrial thickness $\geq$ 4.5 mm.

- To construct mathematical models to evaluate the individual risk of endometrial malignancy in women with postmenopausal bleeding and sonographic endometrial thickness $\geq$ 4.5 mm, using clinical data, sonographic endometrial thickness, grey-scale ultrasound morphology of the endometrium, and power Doppler ultrasound findings.

- To determine whether endometrial volume or power Doppler indices, as measured by 3D ultrasound, can discriminate between benign and malignant endometrium in women with postmenopausal bleeding and sonographic endometrial thickness $\geq$ 4.5 mm, to compare the diagnostic performance of these indices with that of endometrial thickness measurements using 2D ultrasound, and to determine whether 3D power Doppler indices provide any additional information to endometrial thickness or volume.

- To determine which SCSH findings best discriminate between benign and malignant endometrium in women with postmenopausal bleeding and sonographic endometrial thickness $\geq$ 4.5 mm, and to compare the interobserver reproducibility of 2D SCSH and 3D SCSH and the agreement of these techniques with hysteroscopy.

Subjects and methods

Postmenopausal Bleeding Clinic

Since 2002, women seeking care for postmenopausal bleeding at the Department of Obstetrics and Gynaecology, Skåne University Hospital, Malmö, are managed at a postmenopausal bleeding clinic run by two gynaecologists specializing in gynaecological ultrasound. The first procedure is a cervical smear for cytological analysis. Patients then undergo a transvaginal ultrasound examination, during which the endometrial thickness is measured. If the endometrial thickness is ≤ 4.4 mm, the woman is given no further treatment, but asked to come back if bleeding recurs. If the thickness of the endometrium is ≥ 4.5 mm SCSH is performed. If no focal lesions are seen in the uterine cavity at SCSH, an endometrial sample is taken using Endorette, or dilatation and curettage (D&C) is suggested. If focal lesions are seen at SCSH, or if SCSH fails, diagnostic hysteroscopy with hysteroscopic resection of focal lesions/endometrium is recommended.

Patients

Between November 2002 and June 2009, 729 women with postmenopausal bleeding were examined at the Postmenopausal Bleeding Clinic at the Department of Obstetrics and Gynaecology, Skåne University Hospital, Malmö, Sweden. Of these, 371 were eligible for inclusion in the studies (Figure 1).

Women with an endometrial thickness ≥ 4.5 mm and no fluid in the uterine cavity at the transvaginal scan were eligible for inclusion in all studies. Different numbers of women are included in the different studies because of changes in the ultrasound systems, cervical stenosis making it impossible to perform SCSH, or the absence of reliable histological diagnosis. Some patients are included in more than one study, see Table 1.

729 women with postmenopausal bleeding

Endometrium ≤ 4.4 mm (n = 290)

Endometrium not measurable (n = 46)

Fluid in the uterine cavity (n = 20)

Transvaginal examination not possible (n = 2)

371 women eligible for inclusion
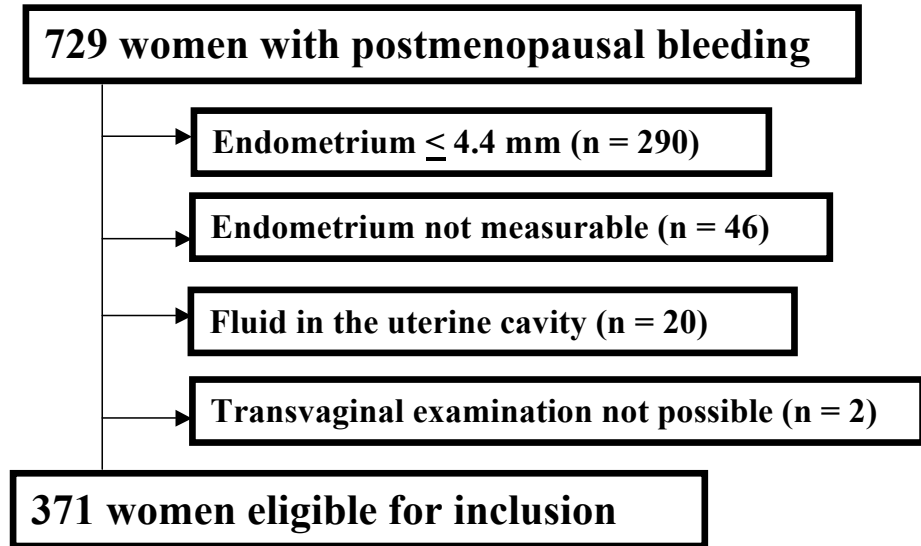
Figure 1. Patient flow chart.

Table 1. Number of patients included in the four studies and overlap between the study populations.

	Study I (n = 120)	Study II (n = 261)	Study III (n = 62)	Study IV (n = 84)
Study I (n = 120)	**120**	98	44	57
Study II (n = 261)		**261**	56	73
Study III (n = 62)			**62**	37
Study IV (n = 84)				**84**

Patient history

A structured history was obtained using a standardized research protocol regarding age at menopause, parity, HRT, weight, height, hypertension, diabetes and current use of anticoagulants. A woman was considered to be postmenopausal if she reported the absence of menstruation for at least 1 year after the age of 40, provided that the amenorrhoea was not explained by pregnancy, medication or disease. Postmenopausal bleeding was defined as any vaginal bleeding in a postmenopausal woman not on HRT, or unscheduled vaginal bleeding in a postmenopausal woman on HRT.

Ultrasound examinations and analysis of ultrasound images and videotapes

Transvaginal sonography was carried out using a Sequoia 512 ultrasound system (Siemens Medical Solutions Inc., Ultrasound Division, Mountain View, CA, USA) equipped with a 5-8 MHz transvaginal transducer (Studies I & II). The women were then examined using a GE Voluson 730 Expert ultrasound system (GE Healthcare, Zipf, Austria) with a 5-9 MHz transvaginal transducer (Studies III & IV).

In Study I, conventional grey-scale ultrasound examination of the uterus was performed (Figure 2), and power Doppler ultrasound examination was carried out using predetermined, standardized settings (Figure 3). The examinations were videotaped for later analysis. Two examiners reviewed all the videotapes, and assessed the endometrial morphology and vascularity using a fixed study protocol. This protocol included predetermined definitions of endometrial morphological characteristics and blood vessel characteristics. The analysis of grey-scale endometrial morphology included visual evaluation of the following: the presence of bright lines separating the endometrial echo from the myometrium, the presence of a middle echo, regularity of the endometrial–myometrial border, internal endometrial structure and homogeneity of endometrial echogenicity (homogeneous, heterogeneous, impossible to evaluate).

Analysis of the videotaped power Doppler ultrasound examinations included visual evaluation of the following: number of blood vessels crossing the myometrial–endometrial border, the size of the blood vessels, the regularity of vessel branching, the presence of large areas of colour, i.e., 'colour splashes', and the presence of areas of densely packed blood vessels. In addition, endometrial vascularity was classified as multiple vascular
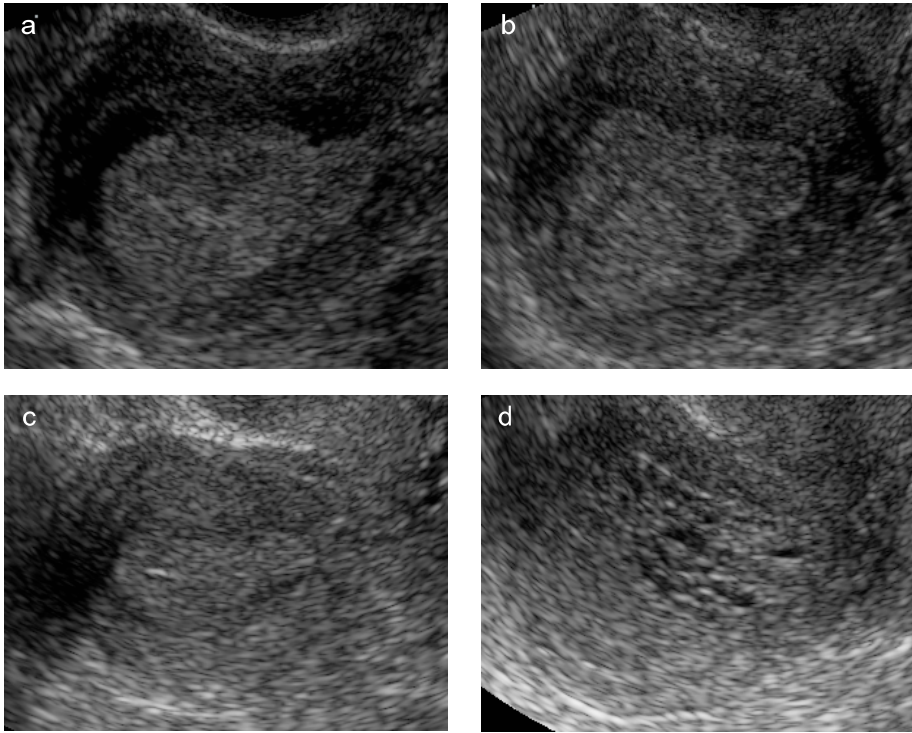
Figure 2. Grey-scale ultrasound examination of the endometrium, showing (a) heterogeneous echogenicity (histopathological diagnosis: adenocarcinoma); (b) heterogeneous echogenicity and irregular endometrial-myometrial border (histopathological diagnosis: adenocarcinoma), (c) homogeneous echogenicity and middle echo (histopathological diagnosis: benign estrogen-influenced endometrium); (d) cystic structure and hyperechoic lines (histopathological diagnosis: polyp).
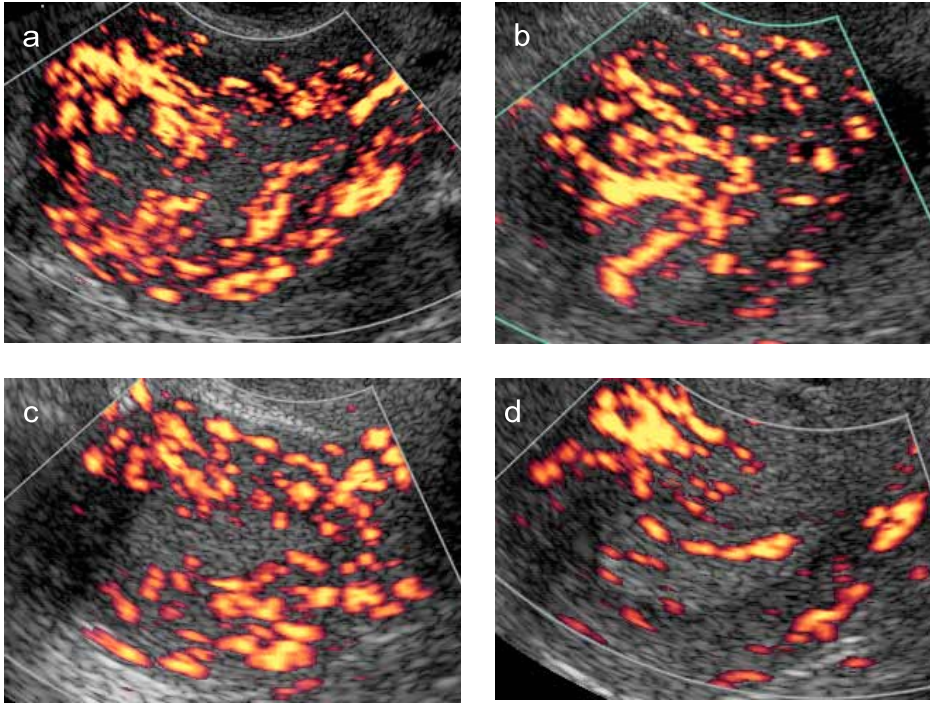
Figure 3. Power Doppler ultrasound examination of the endometrium, showing (a) multiple, densely packed endometrial blood vessels (histopathological diagnosis: adenocarcinoma); (b) multiple, irregularly branching endometrial vessels (histopathological diagnosis: adenocarcinoma), (c) 'scattered' endometrial vessels (histopathological diagnosis: benign estrogen-influenced endometrium; (d) one dominant endometrial vessel (histopathological diagnosis: polyp).

pattern (A), single vascular pattern (B), or scattered vascular pattern (C), as proposed by Alcazar et al. (2003).

The two examiners evaluated the videotapes independently, and the classification subsequently agreed upon by the two examiners was used for statistical analysis. The interobserver reproducibility of the evaluation of endometrial morphology and vascularity was also evaluated.

In Study II, the image of the most vascularized section through the endometrium at power Doppler ultrasound examination, as estimated subjectively, was frozen, and the endometrium was outlined using the trace function of the ultrasound system. The image was then postprocessed, i.e. the grey-scale echoes were removed, so that the coloured power Doppler pixels were shown on a black background (Figure 4). The postprocessed images were saved and transferred to an offline computer where they were analysed by a single examiner using dedicated software (developed using MATLAB 6.0, Marret et al., 2005) to quantify the colour content of the endometrial scan - the percentage vascularized area, Vascularity index (Figure 5). The ultrasound examiners also graded the colour content in the endometrium on a VAS from 0 to 100 during live scanning.

To determine the intra-observer repeatability and the reliability of the calculations of Vascularity index, images from 20 patients were analysed twice by one examiner.

To determine the intra-observer repeatability and reliability and the interobserver reproducibility and reliability of the VAS score, images from 30 patients were used. The 30 images were analysed twice with regard to the VAS score by two independent observers blinded to each others' results.

In Study III, using a GE Voluson 730 Expert ultrasound system, the system was switched to power Doppler mode. Identical pre-installed settings were used for all patients. The 3D mode was activated, the corpus uteri (without the cervix) was centralized within the 3D ultrasound sector on the screen so that it filled the whole 3D ultrasound sector, and a volume of the corpus uteri was acquired. The volume was saved for later analysis. The stored 3D ultrasound volumes of the corpus uteri were analysed by one observer. The endometrial volume and the three power Doppler indices (VI, FI and VFI) were calculated in the endometrium and in a 2 mm layer (i.e. the subendometrium) surrounding the endometrium using the VOCAL software (Figure 6, 7). Rotation steps of 30 degrees were

used. The A plane (sagittal view of the uterus) was rotated around the Y-axis with all measurements being conducted on plane A. To determine the intra-observer repeatability and reliability, the volumes of 30 patients were analysed by one observer twice.

In Study IV, after performing the conventional grey-scale ultrasound examination, SCSH was carried out. The 2D SCSH images were videotaped (Figure 8). Thereafter, a 3D ultrasound volume of the corpus uteri containing saline was acquired (Figure 9). After SCSH, an endometrial sample was taken using the Endorette endometrial sampling device. Unless the Endorette sample revealed a malignancy, women with focal lesions in the uterine cavity were recommended hysteroscopic resection of the focal lesion(s) under general anaesthesia. Hysteroscopy was performed by any of six gynaecologists specifically trained to perform this procedure. Women with no focal lesions were recommended D&C. After completion of the study, the videotapes and 3D ultrasound volumes were independently analysed by two examiners, the 3D ultrasound volumes being analysed using the VOCAL software, the Z technique (Abuhamad et al., 2006) and the orientation guidelines described by Merz et al. (2007). A standardized research protocol was followed when evaluating the ultrasound findings. For women who underwent hysteroscopy, a similar standardized protocol was filled in by the hysteroscopist immediately after the procedure. The hysteroscopist had no knowledge of the ultrasound findings when performing the hysteroscopy. The presence, number and surface regularity of focal lesions were evaluated by the ultrasound examiner and the hysteroscopist. In the absence of focal lesions, the surface regularity of the endometrium was evaluated by the ultrasound examiner only.

Agreement with regard to the location of the focal lesions was classified as complete agreement, partial agreement, disagreement and not possible to assess. After having independently assessed all the videotapes and 3D ultrasound volumes, the two ultrasound observers reviewed their findings together. If there was disagreement in any of the variables they went back to the videotapes/volumes and reviewed them together to arrive at a consensus. Their consensus opinion of the uterine cavity was compared with the final histopathological diagnosis and with the hysteroscopy findings.
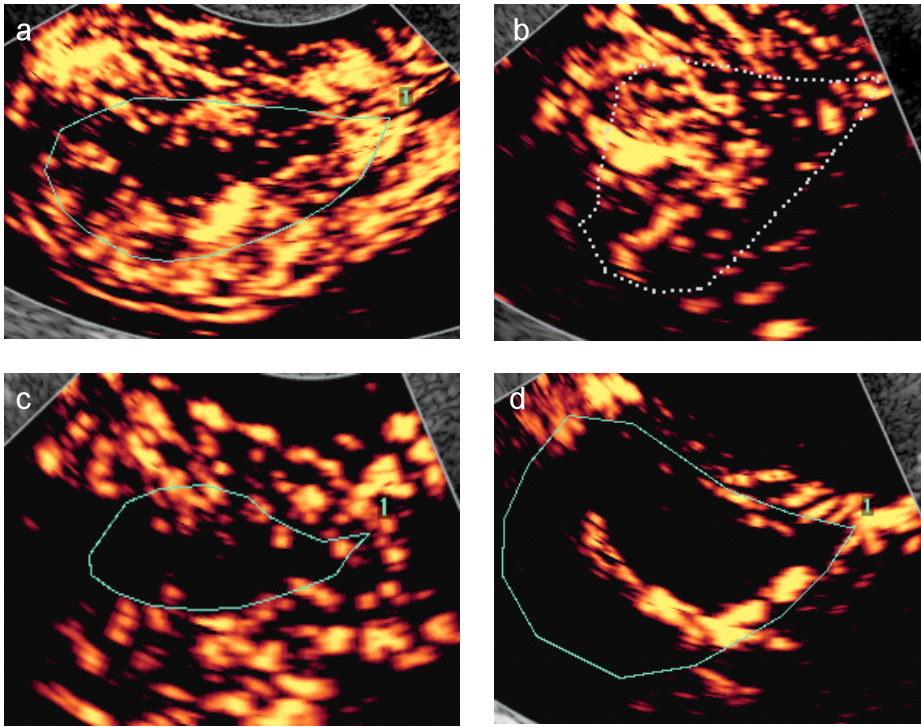
Figure 4. Postprocessed endometrial images showing (a) multiple, densely packed endometrial blood vessels (histopathological diagnosis: adenocarcinoma); (b) multiple, irregularly branching endometrial vessels (histopathological diagnosis: adenocarcinoma), (c) 'scattered' endometrial vessels (histopathological diagnosis: benign estrogen-influenced endometrium; (d) one dominant endometrial vessel (histopathological diagnosis: polyp). The dotted or green lines in the images outline the endometrium.
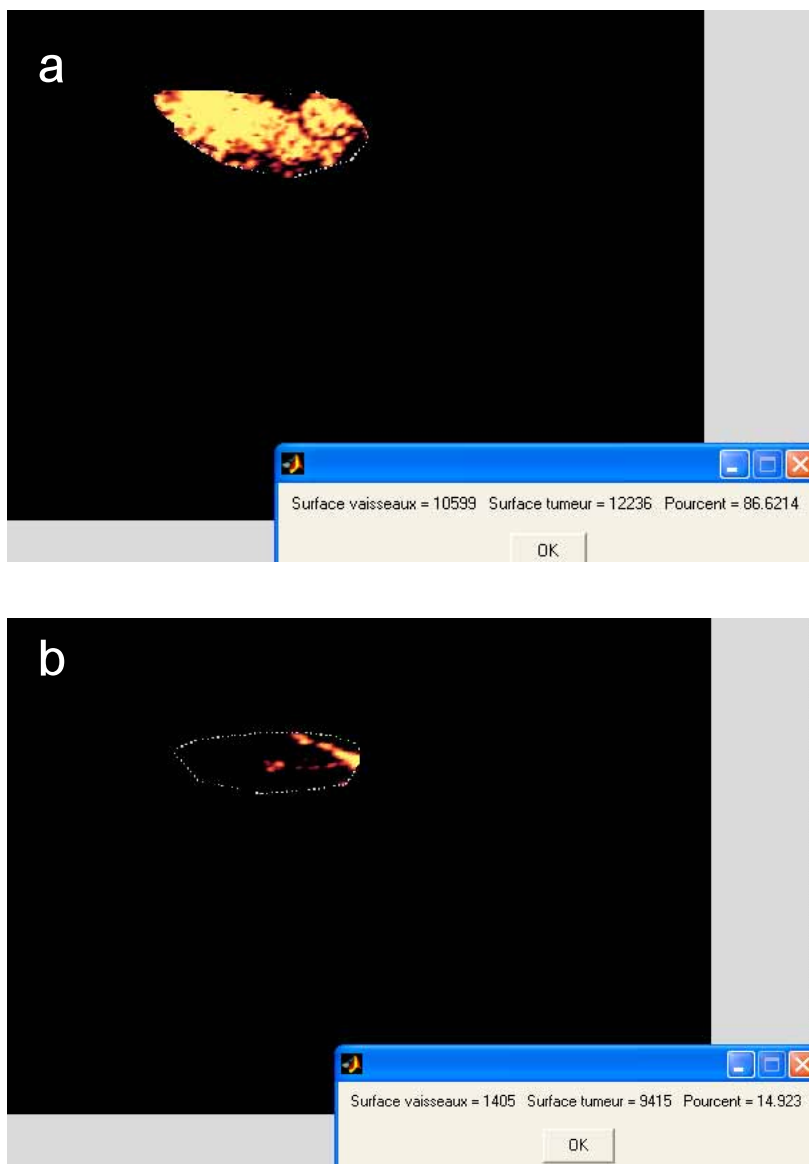
Figure 5. Calculation using the MATLAB software of the percentage of vascularized area (Vascularity index) (a) Vascularity index 86% (histopathological diagnosis: adenocarcinoma); (b) Vascularity index 15% (histopathological diagnosis: polyp).

Figure 6. Calculation using VOCAL software of endometrial volume (a) and power Doppler indices (VI, FI, VFI) (b) in the endometrium (histopathological diagnosis: adenocarcinoma).

Figure 7. Calculation using VOCAL software of the subendometrial volume (a) and power Doppler indices (VI, FI, VFI) (b) in the subendometrium (histopathological diagnosis: adenocarcinoma).
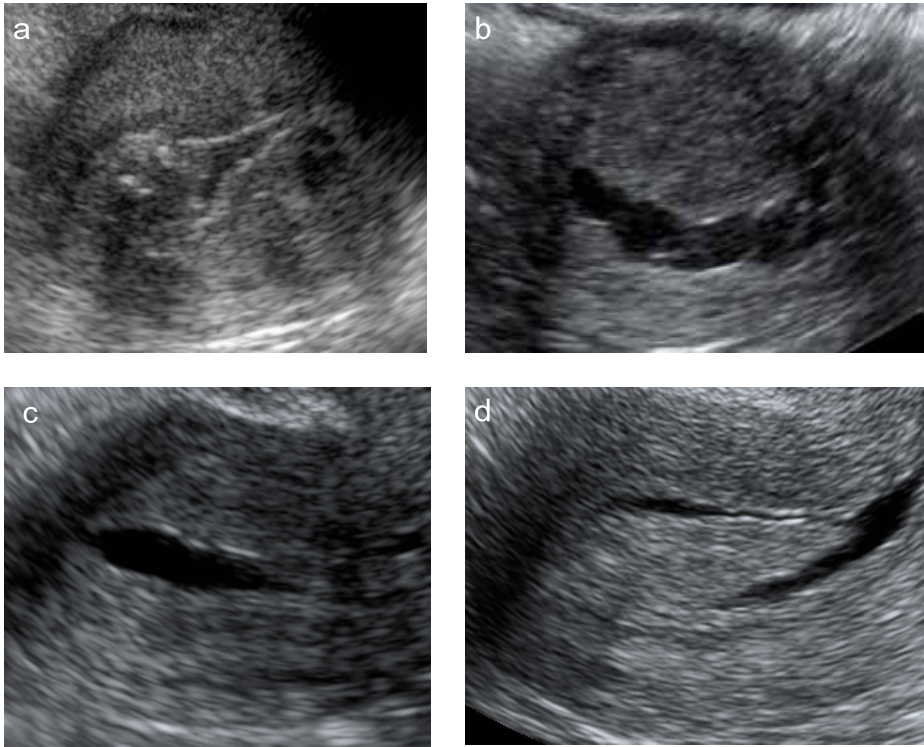
Figure 8. 2D SCSH ultrasound images showing (a) irregular surface of focal lesion, distension difficulties (histopathological diagnosis: adenocarcinoma); (b) irregular surface of focal lesion, heterogeneous echogenicity (histopathological diagnosis: adenocarcinoma); (c) smooth endometrial cavity (histopathological diagnosis: benign estrogen-influenced endometrium); (d) smooth focal lesion (histopathological diagnosis: polyp).
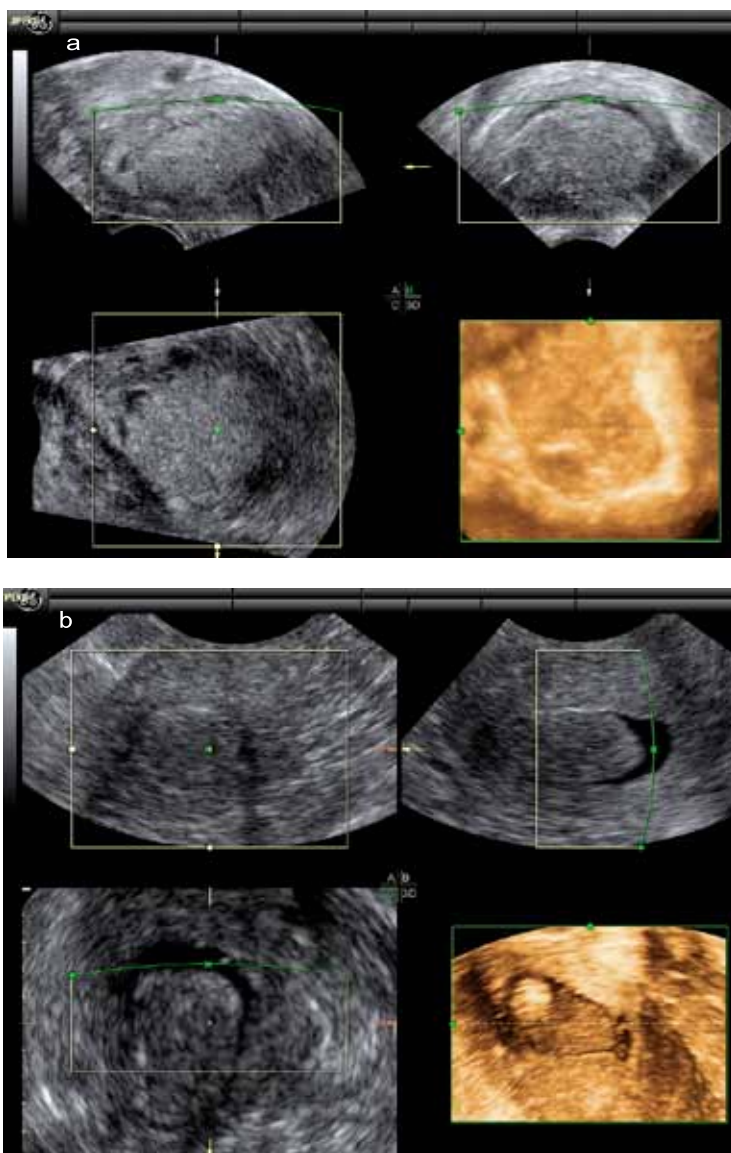
Figure 9. 3D SCSH ultrasound images, (a) histopathological diagnosis: adenocarcinoma;
(b) histopathological diagnosis: polyp.

The reference standard

Clinical and ultrasound findings were compared with the final histological diagnosis of the endometrium, the final histological diagnosis being obtained by D&C (when there were no focal lesions in the uterine cavity at SCSH) or hysteroscopic resection (when focal lesions were seen in the uterine cavity at SCSH, or if SCSH failed) or by hysterectomy. Staging of malignant tumors was done by the responsible onco-gynaecologist in accordance with the classification system recommended by the International Federation of Gynaecology and Obstetrics (Shepherd, 1989).

Statistical methods

Statistical calculations were performed using the Statistical Package for the Social Sciences (SPSS Inc., Chicago, Illinois, USA, versions 12.02 and 16.0), and the statistical software StatXact (Cytel Inc., Cambridge, MA, USA, version 4). All the studies were observational, and the following statistical methods were used in the studies.

The *Mann-Whitney test* was used to determine the statistical significance of a difference in unpaired continuous data (as the data were not normally distributed) and *Fisher's exact test* was used to determine that of a difference in unpaired discrete data. A two-tailed p-value <0.05 was considered to be statistically significant (Studies I & IV).

The *McNemar test* was used to determine if differences in agreement between 2D and 3D hydrosonography were statistically significant; a two-tailed p-value <0.05 being considered statistically significant (Study IV).

Possible relationships between single ultrasound variables, clinical variables and endometrial malignancy were determined using *univariate logistic regression* with the likelihood ratio test; a two-tailed p-value <0.05 being considered statistically significant (Studies I-IV). *Multivariate logistic regression* analysis with the likelihood ratio test was used to create mathematical models to calculate the risk of endometrial malignancy (Studies I-IV).

Receiver operating characteristic (ROC) curves were drawn for single predictive clinical and ultrasound variables and for the multivariate logistic regression models. The area under the ROC curve (AUC) and the 95% confidence interval (CI) of this area were calculated. If the lower limit of the CI for the AUC was >0.5, the variable/model was

considered to have discriminatory potential. The ROC curves were also used to determine the mathematically best cut-off value of continuous single variables for predicting malignancy, as well as for logistic regression models; the mathematically best cut-off value being defined as that corresponding to the point on the ROC curve farthest from the reference line. The sensitivity, specificity, positive and negative likelihood ratios (LR+ and LR-) of the mathematically best cut-off value and their 95% CIs were then calculated (Jaeschke et al., 1994). The best diagnostic test was defined as that with the largest AUC (Studies I-IV).

Percentage agreement between observers and between SCSH and hysteroscopy was determined (Study IV) (Kundel et al, 2003), and *Cohen's Kappa* coefficient was calculated to assess interobserver agreement; values of 0.81-1 being taken to indicate almost perfect agreement, values of 0.61-0.8 good agreement, and values of 0.41-0.6 moderate agreement (Studies I & IV) (Kundel et al, 2003).

Intra-observer repeatability and interobserver reproducibility were estimated as the difference between the results of two measurements (Bartlett & Frost., 2008). The differences between the measured values were plotted against the mean of the two measurements (Bland-Altman plots, Bland & Altman, 1986) to assess the relationship between the differences and the magnitude of the measurements. Systematic bias between two analyses was estimated by calculating the 95% CI of the mean difference (mean difference ± 2 SE, standard error of the mean). If zero lay within this interval it was assumed that there was no bias between the two measurements. Intra-observer repeatability was expressed as the *repeatability coefficient*; the absolute difference between two measurements on a subject being expected to differ by no more than the repeatability coefficient on 95% of occasions (Studies II & III).

Intra- and interobserver reliability were determined by calculating the *intra-class correlation coefficient (ICC)* using analysis of variance (two-way random model – absolute agreement, which allows generalization of the results to a population of observers). The ICC indicates the proportion of the total variance in measured results that can be explained by differences between the individuals examined. A high ICC indicates that the measurements can be used to discriminate between individuals. The more variable the population investigated, the greater the ICC (Studies II & III).

Results and discussion

Study I

Of the 120 women included in Study I, 90 (75%) had benign and 30 (25%) had malignant endometrium. The grey-scale ultrasound morphology variable that best predicted malignancy was *heterogeneous endometrial echogenicity* (AUC 0.83, sensitivity 73%, specificity 92%, LR+ 9.4, LR- 0.3), and the power Doppler ultrasound variable that best predicted malignancy was *irregular branching of endometrial blood vessels* (AUC 0.77, sensitivity 60%, specificity 94%, LR+ 10.8, LR- 0.4).

None of the clinical variables entered a logistic regression model to predict malignancy. The best logistic regression model with only two ultrasound variables was a model including *endometrial thickness* and *heterogeneous echogenicity of the endometrium* (AUC 0.91). The diagnostic performance improved marginally when power Doppler information (*areas of densely packed blood vessels* or *irregular branching of endometrial blood vessels*) was added to this model (AUC 0.92).

Interobserver agreement in the evaluation of endometrial vascularity was superior to that in the evaluation of grey-scale ultrasound images (Cohen's Kappa 0.49-0.78 and 0.50-0.66, respectively).

Study II

Of the 261 women included in Study II, 198 (76%) had benign and 63 (24%) had malignant endometrium. According to the AUC, age, the use of HRT, sonographic endometrial thickness, Vascularity index and the VAS score had a potential to discriminate between benign and malignant endometrium.

Models including only clinical variables showed poor diagnostic performance, the best of these having an AUC of 0.74. Adding endometrial thickness to the clinical variables substantially improved the diagnostic performance; the best model including clinical variables and endometrial thickness had an AUC of 0.82. Adding both endometrial thickness and power Doppler ultrasound results to the clinical variables improved model

performance even further. The model with the largest AUC (0.91) and the highest specificity at 90% sensitivity included the variables *endometrial thickness*, *Vascularity index*, *age*, and *use of HRT*. Using a risk cut-off of 0.11 (i.e., 11% risk) to indicate malignancy, this model had a sensitivity of 90.5%, a specificity of 71%, LR+ 3.14 and LR- 0.13.

Intra- and interobserver reliability in the evaluation of the VAS score was good (all ICC values >0.95), and intra-observer reliability in the analysis of Vascularity index was excellent (ICC = 1.00).

Study III

Of the 62 women included in Study III, 49 (79%) had benign endometrium and 13 (21%) had malignant endometrium. The endometrial volume was larger and the flow indices were higher in both the endometrium and in the subendometrium in patients with malignant endometrium than in those with benign endometrium, but there was substantial overlap between the two categories. The best variables for discriminating between benign and malignant endometrium were *endometrial thickness*, and *endometrial VI* and *VFI*, all having AUC of 0.82.

The best logistic regression model for predicting malignancy in Study III was that including the variables *endometrial thickness* and *VI in the subendometrium* with AUC of 0.86. Using the mathematically optimal risk cut-off value (0.22), the model correctly classified seven more benign cases, but two fewer malignant cases, than the best endometrial thickness cut-off (11.8 mm). Models including endometrial volume and flow indices performed less well than endometrial thickness alone (AUC 0.79 vs. 0.82).

Intra-observer reliability was very high for all ultrasound variables (all ICC values >0.95).

Study IV

Of the 84 women included in Study IV, 70 (83%) had benign and 14 (17%) had malignant endometrium. The ultrasound variable that best discriminated between benignity and malignancy for both 2D and 3D SCSH was the presence of *at least one focal lesion with an irregular surface* (for 2D SCSH the AUC was 0.84, sensitivity 71%, specificity 97%,

LR+ 25 and LR- 0.3; for 3D SCSH the AUC was 0.70, sensitivity 43%, specificity 97%, LR+ 15 and LR- 0.6).

Interobserver agreement with regard to the presence of focal lesions and the presence of irregular focal lesions was similar for 2D and 3D SCSH (88-98%). Interobserver agreement for the number and localization of focal lesions was also similar for 2D and 3D SCSH (63-76% for the number of lesions, and 70-76% for localization), but it was poorer than that for the presence of focal lesions.

The agreement between 2D SCSH and hysteroscopy with regard to the presence of focal lesions, the presence of focal lesions with an irregular surface, and the localization of focal lesions was similar to that between 3D SCSH and hysteroscopy (94% vs. 93%; 74% vs. 76%; 66% vs. 63%).

Comparison of the results

Different samples of patients were included in the four studies, so it is not possible to draw definitive conclusions about which methods best predict endometrial malignancy, but it is still of interest to compare the results. In all the studies women with malignant endometrium were older, they had thicker endometrium and fewer used HRT than those with benign endometrium (Table 2).

Table 2. Age, hormonal replacement therapy (HRT) and endometrial thickness (ET) of the patients included in all the studies.

Study	Age (years, median)			HRT (%)			ET (mm, median)		
	Malignant	Benign	p	Malignant	Benign	p	Malignant	Benign	p
I	73	63	.023	10	16	.560	17.6	10.2	.0005
II	74	65	.0005	14	35	.001	20.8	10.2	.0005
III	73	63	.085	23	43	.181	16.2	9.6	.001
IV	68	64	.334	7	18	.028	16.1	10.3	.001

p-value, univariate logistic regression analysis (Studies II & III), Mann-Whitney and Fisher's exact test (Studies I & IV)

The results of all four studies showed that ultrasound imaging can discriminate between benign and malignant endometrium, and predict endometrial malignancy in women with postmenopausal bleeding and sonographic endometrial thickness $\geq$ 4.5 mm. Endometrial thickness was found to be a fairly good predictor of malignancy in all studies (Table 3), and the risk of malignancy increased with increasing sonographic endometrial thickness. This is in agreement with the results of others (Epstein et al., 2002, Randelzhofer et al., 2002, Weber et al., 1998).

Table 3. Diagnostic performance of endometrial thickness in the prediction of endometrial malignancy in the four studies.

Study	AUC	Cut-off value (mm)	Sensitivity (%)	Specificity (%)	p
I	0.80	15.0	73	77	0.0005
II	0.80	15.3	70	81	0.0005
III	0.82	11.8	85	71	0.001
IV	0.78	12.5	93	69	0.013

AUC, area under the ROC curve; Cut-off value, mathematically optimal value corresponding to the point on the ROC curve farthest from the reference line; p-value, univariate logistic regression analysis.

The diagnostic performance of the ultrasound variables with the best diagnostic performance with regard to the prediction endometrial malignancy in each study is given in Table 4.

The diagnostic performance of ultrasound variables other than endometrial thickness (Table 4) was similar to that of sonographic endometrial thickness (Table 3). Only the presence of a focal lesion with an irregular surface during SCSH and the VAS score for the colour content of the endometrial scan using power Doppler ultrasound were slightly better (AUC 0.84 vs. 0.82). Determining the presence of focal lesions with irregular surface during SCSH is a valuable procedure in women with postmenopausal bleeding and thick endometrium. Irregular surface of the endometrium has also been found to be a sign

Table 4. Diagnostic performance of ultrasound variables for the prediction of endometrial malignancy.

Variable	Study	AUC	Sensitivity (%)	Specificity (%)	p
Focal lesion with irregular surface	IV	0.84	71	97	0.0005
VAS score of colour content	II	0.84	67	95	0.0005
Heterogeneous endometrium	I	0.83	73	92	0.0005
Vascularity index	II	0.82	70	88	0.0005
VI	III	0.82	69	84	0.008

AUC, area under the ROC curve; VI, vascularization index; p value, univariate logistic regression analysis

of endometrial malignancy in other studies (Epstein & Valentin, 2006, Bronz et al., 1997, Laifer-Narin et al., 1999, Parsons, 2002). However, in agreement with others, we found that patients with postmenopausal bleeding often have cervical stenosis (17% (29/170), Study IV), making it impossible to perform SCSH (De Silva et al., 1007, Epstein et al., 2001, De Kroon et al., 2003). SCSH is usually not a painful procedure (Van den Bosch et al., 2008), but postmenopausal women experience pain during SCSH more often than premenopausal women (73% (22/30) vs. 36% (29/81), p<0.0005 (unpublished results from my research at Vilnius University Hospital, Lithuania, where patients filled in questionnaires to evaluate their pain). This could be explained by postmenopausal women more often having a narrow cervical canal making insertion of the plastic catheter into the uterus difficult, or other postmenopausal changes in the uterus.

The subjective evaluation of the colour content in the endometrium by experienced examiners using power Doppler ultrasound also seems to be a valuable method for predicting endometrial malignancy (Table 4). Objective quantification of the colour content of the endometrial scan (i.e. the percentage vascularized endometrial area,

Vascularity index) had a similar diagnostic performance in predicting endometrial malignancy to the VAS score (AUC 0.82 vs. 0.84), and the same predicative performance as the 3D ultrasound variable VI in the endometrium (Table 4). The diagnostic performance of the quantification of the colour content in 3D power Doppler ultrasound images of the endometrium (VI) was not superior to evaluation of the colour content in 2D ultrasound images of the endometrium (VAS, Vascularity index) with regard to discrimination between benign and malignant endometrium. The diagnostic performance of 3D power Doppler ultrasound for the prediction of endometrial cancer has been found to vary between studies (Alcazar et al, 2007, Odeh et al, 2007, Yaman et al, 2002).

Heterogeneous endometrial echogenicity in grey-scale ultrasound examination (the endometrium has a 'moth-eaten' appearance) had the highest sensitivity (73%) with regard to endometrial cancer, and only a slightly smaller AUC (0.83 vs. 0.84) than the other ultrasound variables. It has also been found in other studies that heterogeneous endometrial structure is a sign of endometrial cancer (Epstein & Valentin, 2006, Sheth et al., 1993).

The logistic regression models developed in Studies I, II and III had better diagnostic performance than endometrial thickness alone, and than any other single ultrasound variable. The models with the best performance are listed in Table 5.

The logistic regression model including the variables *endometrial thickness* and *heterogeneous endometrial echogenicity* showed the highest sensitivity in predicting malignancy (93%). Adding power Doppler information (*irregular branching of vessels*) to this model improved the diagnostic performance marginally (AUC 0.92 vs. 0.91), but the sensitivity was reduced to 87%, resulting in more endometrial malignancies being missed (four vs. two) using the model including power Doppler information. The diagnostic performance of the model including endometrial thickness and the 3D power Doppler variable VI in the subendometrium was not superior to that of the model with 2D power Doppler information (0.86 vs. 0.91, Table 5).

Table 5. Diagnostic performance of the logistic regression models for the prediction of endometrial malignancy.

Model	Study	AUC	Sensitivity (%)	Specificity (%)
ET, heterogeneous endometrium, irregularly branching of vessels	I	0.92	87	83
ET, heterogeneous endometrium	I	0.91	93	79
Age, HRT, ET, Vascularity index	II	0.91	89	77
ET, VI in the subendometrium	III	0.86	69	86

AUC, area under the ROC curve; ET, endometrial thickness; VI, vascularization index; HRT, hormonal replacement therapy; subendometrium, 2 mm layer surrounding the endometrium

Study II clearly demonstrated that the diagnostic performance of logistic regression models for the prediction of endometrial cancer improves when endometrial thickness and power Doppler variables are added to clinical variables. Models including only clinical variables showed poor diagnostic performance. The ROC curves are shown in Figure 9. In agreement with the results of this thesis, Opmeer et al., (2007) also found that adding information on endometrial thickness to clinical information substantially increased the diagnostic performance of risk calculation models. The AUC in their model increased from 0.76 to 0.90 when they added endometrial thickness to clinical information, but their study included also women with thin endometrium.
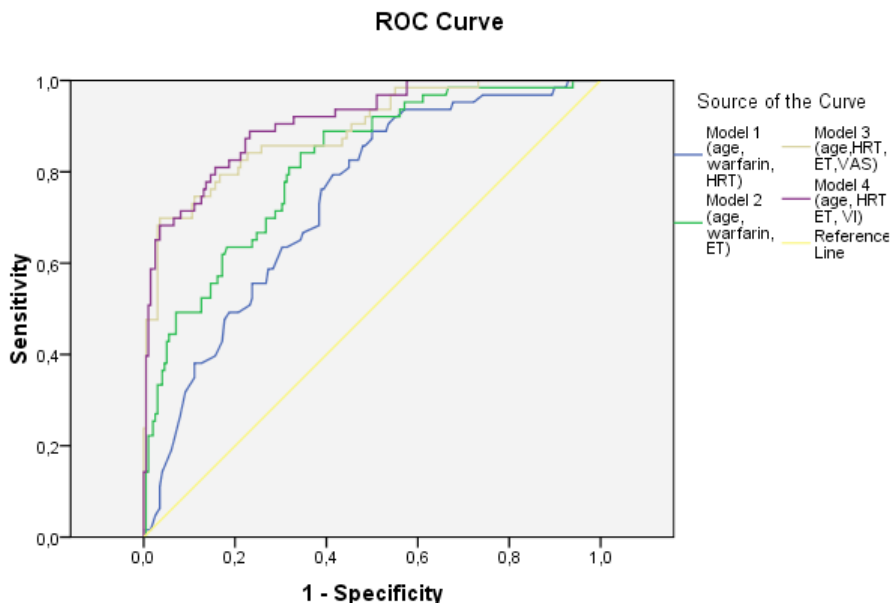
Figure 9. ROC curves of models for estimating the risk of endometrial malignancy.
Model 1 (age, use of warfarin, use of HRT), AUC 0.74; Model 2 (age, use of warfarin, endometrial thickness), AUC 0.82; Model 3 (age, use of HRT, endometrial thickness, VAS score), AUC 0.89; Model 4 (age, use of HRT, endometrial thickness, Vascularity index), AUC 0.91; The yellow reference line represents the ROC curve for a useless test.

One strength of this work is that only women with a reliable histological diagnosis are included. Cases where there was insufficient endometrial material for histological diagnosis, and cases where the final diagnosis was based on the Endorette sample only were excluded. Most other studies in which logistic regression models were created to calculate the risk of endometrial malignancy also included women whose diagnosis was based on an endometrial sample taken using an outpatient endometrial sampling device

(Ferrazzi et al., 1996, Randelzhofer et al., 2002, Opmeer et al., 2007). I believe that a well-defined study population with regard to histopathological diagnosis is preferable to a larger, but less well-defined, study population. Women using HRT were included so that the models for the estimation of risk of endometrial cancer would also be applicable to these women. A second strength is that the reproducibility of the assessment of different ultrasound variables was determined, and the results showed that the reproducibility is good enough for clinical use.

A limitation of this work is that information on all the ultrasound variables used to predict malignancy was not available for all the women. One reason for this is that SCSH failed in many women, most often because of cervical stenosis. Another reason is the upgrading of the 3D ultrasound system used, resulting in the values of VI, FI and VFI after upgrading being different from those before upgrading. Because, information on all ultrasound variables was not available for all the women a reliable comparison of the diagnostic performance of all the variables and models is not possible. However, it was possible to compare the diagnostic performance of the variables in Study I with those in Study II on a sample of 98 women (27 with endometrial cancer) included in both studies. This comparison showed that as single variables the grey-scale variable *heterogeneous endometrial echogenicity* and the power Doppler ultrasound variables *VAS score* and *Vascularity index* best predicted endometrial malignancy (according to the AUC). Another limitation of this work is that the results are only applicable to ultrasound examinations carried out using the same power Doppler ultrasound system and transducer as were used in the current studies. This applies to all studies where colour and power Doppler variables are involved, since the colour content of a colour or power Doppler scan depends heavily on the Doppler sensitivity of the ultrasound system used.

It is important to remember that all four studies in this thesis included only women at high risk of endometrial cancer, i.e., women with postmenopausal bleeding and endometrial thickness ≥4.5 mm. Moreover, they included only women with a reliable histological diagnosis. Therefore, it is difficult to compare the results presented here with those of others. Other studies included women with thin endometrium and with a final diagnosis based only on endometrial biopsy (Ferrazi et al., 1996, Randelzhofer et al., 2002). One study even included women with no histological diagnosis (i.e. women not known to have

been diagnosed of endometrial cancer within 6 months of the ultrasound examination were classified as having a benign endometrium) (Opmeer et al., 2007).

I believe that the best logistic regression models presented in this thesis could be used in clinical practice for the calculation of individual risk of malignancy in order to tailor patient management. For example, in a woman with an estimated low risk of endometrial cancer at high operative risk, it may be appropriate to refrain from trying to obtain an endometrial sample, at least if cervical stenosis or other factors make it impossible to obtain an endometrial sample using an outpatient sampling device. On the other hand, a high risk of malignancy is an indication for the use of a reliable diagnostic procedure. The logistic regression models presented in this thesis must, however, be externally validated.

Conclusions

Based on the work presented in this thesis, the following conclusions can be drawn regarding women with postmenopausal bleeding and endometrial thickness $\geq$ 4.5 mm:

- Endometrial thickness is a powerful predictor of endometrial malignancy, even in a high-risk group of women with postmenopausal bleeding and endometrial thickness $\geq$ 4.5 mm.
- Endometrial malignancy must be suspected when a thick endometrium with heterogeneous echogenicity is found at grey-scale ultrasound examination.
- The finding of a high colour content in the endometrium, by subjective (VAS scale) or objective (Vascularity index) quantification of the colour content using power Doppler ultrasound, increases the risk of endometrial malignancy.
- The finding of one or more focal lesions with an irregular surface during SCSH increases the risk of endometrial malignancy.
- The diagnostic performance of logistic regression models to calculate the individual risk of endometrial malignancy improves substantially if ultrasound information is added to clinical variables.
- 3D ultrasound is not superior to 2D ultrasound for the prediction of endometrial malignancy.

Acknowledgements

I would like to thank all who have been in my life during my research and have contributed to this work:

Professor *Lil Valentin*, my supervisor, for opening to me the wonderful world of science, for sharing your endless knowledge with me, for your inspiration and belief in me, for our scientific talks. You have taught me everything I know in the field of research and have given to me the opportunity to participate in the scientific life.

Associate professor *Povilas Sladkevicius*, my co-supervisor, for introducing me to Professor Lil Valentin and so opening to me the way for research in Sweden, for unselfishly helping me with both scientific and practical things, for reading thoroughly my thesis and for helping me to organize the printing of this book.

Dr *Gunilla Bodelsson*, Director of the Department of Obstetrics and Gynaecology in Malmö, for friendly and scientific atmosphere in the clinic and for an always supporting smile.

Associate professor *Christer Borgfeldt* and Dr *Louise Thunell*, my 'half-time control' examiners, for reading my papers and for constructive critique.

Dana Sladkeviciene, for your hospitality, soft sofa (when it was necessary), practical help and tasteful dinners.

My friends *Anna Sokalska* (for your friendship and heartiness during our first days in Sweden), *Ligita Jokubkiene* (for your practical advises and for giving me the opportunity to live in your flat), *Linas Rovas* (for your heartiness and life advises), *Olga Vikhareva Osser* (for our warm and simple Russian talks).

References

Abuhamad AZ, Singleton S, Zhao Y, Bocca S. The Z technique: an easy approach to the display of the mid-coronal plane of the uterus in volume sonography. *J Ultrasound Med*. 2006 May;25(5):607-12.

Abulafia O, Sherer DM. Angiogenesis of the endometrium. *Obstet Gynecol.* 1999 Jul;94(1):148-53.

Alcazar JL, Castillo G, Minguez JA, Galan MJ. Endometrial blood flow mapping using transvaginal power Doppler sonography in women with postmenopausal bleeding and thickened endometrium. *Ultrasound Obstet Gynecol.* 2003 Jun;21(6):583-8.

Alcazar JL, Errasti T, Zornoza A. Saline infusion sonohysterography in endometrial cancer: assessment of malignant cells dissemination risk. *Acta Obstet Gynecol Scand.* 2000 Apr;79(4):321-2.

Alcazar JL, Galvan R. Three-dimensional power Doppler ultrasound scanning for the prediction of endometrial cancer in women with postmenopausal bleeding and thickened endometrium. *Am J Obstet Gynecol.* 2009 Jan;200(1):44 e1-6.

Amant F, Moerman P, Neven P, Timmerman D, Van Limbergen E, Vergote I. Endometrial cancer. *Lancet.* 2005 Aug 6-12;366(9484):491-505.

Bartlett JW, Frost C. Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. *Ultrasound Obstet Gynecol.* 2008 Apr;31(4):466-75.

Berry E, Lindheim SR, Connor JP, Hartenbach EM, Schink JC, Harter J, et al. Sonohysterography and endometrial cancer: incidence and functional viability of disseminated malignant cells. *Am J Obstet Gynecol.* 2008 Sep;199(3):240 e1-8.

Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986 Feb 8;1(8476):307-10.

Bonilla-Musoles F, Raga F, Osborne NG, Blanes J, Coelho F. Three-dimensional hysterosonography for the study of endometrial tumors: comparison with conventional transvaginal sonography, hysterosalpingography, and hysteroscopy. *Gynecol Oncol.* 1997 May;65(2):245-52.

Bree RL, Bowerman RA, Bohm-Velez M, Benson CB, Doubilet PM, DeDreu S, et al. US evaluation of the uterus in patients with postmenopausal bleeding: A positive effect on diagnostic decision making. *Radiology*. 2000 Jul;216(1):260-4.

Bronz L, Suter T, Rusca T. The value of transvaginal sonography with and without saline instillation in the diagnosis of uterine pathology in pre- and postmenopausal women with abnormal bleeding or suspect sonographic findings. *Ultrasound Obstet Gynecol.* 1997 Jan;9(1):53-8.

Danforth's Obstetrics and Gynecology. Eds. James R., Md. Scott, Ronald S., Md. Gibbs, Beth Y., Md. Karlan, Arthur F., Md. Haney, David N. Danforth. Lippincott Williams & Wilkins Publishers; 9th edition (August 2003). pages 525-535.

De Kroon CD, de Bock GH, Dieben SW, Jansen FW. Saline contrast hysterosonography in abnormal uterine bleeding: a systematic review and meta-analysis. *BJOG.* 2003 Oct;110(10):938-47.

De Silva BY, Stewart K, Steven JD, Sathanandan M. Transvaginal ultrasound measurement of endometrial thickness and endometrial pipelle sampling as an alternative diagnostic procedure to hysteroscopy and dilatation and curettage in the management of post-menopausal bleeding. *J Obstet Gynaecol.* 1997 Jun;17(4):399-402.

Dessole S, Rubattu G, Farina M, Capobianco G, Cherchi PL, Tanda F, et al. Risks and usefulness of sonohysterography in patients with endometrial carcinoma. *Am J Obstet Gynecol.* 2006 Feb;194(2):362-8.

Engelsen IB, Akslen LA, Salvesen HB. Biologic markers in endometrial cancer treatment. *APMIS.* 2009 Oct;117(10):693-707.

Epstein E, Ramirez A, Skoog L, Valentin L. Dilatation and curettage fails to detect most focal lesions in the uterine cavity in women with postmenopausal bleeding. *Acta Obstet Gynecol Scand.* 2001 Dec;80(12):1131-6.

Epstein E, Ramirez A, Skoog L, Valentin L. Transvaginal sonography, saline contrast sonohysterography and hysteroscopy for the investigation of women with postmenopausal bleeding and endometrium > 5 mm. *Ultrasound Obstet Gynecol.* 2001 Aug;18(2):157-62.

Epstein E, Skoog L, Isberg PE, De Smet F, De Moor B, Olofsson PA, et al. An algorithm including results of gray-scale and power Doppler ultrasound examination to predict endometrial malignancy in women with postmenopausal bleeding. *Ultrasound Obstet Gynecol.* 2002 Oct;20(4):370-6.

Epstein E, Skoog L, Valentin L. Comparison of Endorette and dilatation and curettage for sampling of the endometrium in women with postmenopausal bleeding. *Acta Obstet Gynecol Scand.* 2001 Oct;80(10):959-64.

Epstein E, Valentin L. Managing women with post-menopausal bleeding. *Best Pract Res Clin Obstet Gynaecol*. 2004 Feb;18(1):125-43.

Epstein E, Valentin L. Gray-scale ultrasound morphology in the presence or absence of intrauterine fluid and vascularity as assessed by color Doppler for discrimination between benign and malignant endometrium in women with postmenopausal bleeding. *Ultrasound Obstet Gynecol*. 2006 Jul;28(1):89-95.

Exalto N, Stappers C, van Raamsdonk LA, Emanuel MH. Gel instillation sonohysterography: first experience with a new technique. *Fertil Steril*. 2007 Jan;87(1):152-5.

Ferrazzi E, Torri V, Trio D, Zannoni E, Filiberto S, Dordoni D. Sonographic endometrial thickness: a useful test to predict atrophy in patients with postmenopausal bleeding. An Italian multicenter study. *Ultrasound Obstet Gynecol*. 1996 May;7(5):315-21.

Fleischer AC, Kalemeris GC, Machin JE, Entman SS, James AE, Jr. Sonographic depiction of normal and abnormal endometrium with histopathologic correlation. *J Ultrasound Med*. 1986 Aug;5(8):445-52.

Fleischer AC, Shappell HW, Parker LP, Hanemann CW. Color Doppler sonography of endometrial masses. *J Ultrasound Med*. 2002 Aug;21(8):861-5.

Fong K, Causer P, Atri M, Lytwyn A, Kung R. Transvaginal US and hysterosonography in postmenopausal women with breast cancer receiving tamoxifen: correlation with hysteroscopy and pathologic study. *Radiographics*. 2003 Jan-Feb;23(1):137-50.

Goldstein RB, Bree RL, Benson CB, Benacerraf BR, Bloss JD, Carlos R, et al. Evaluation of the woman with postmenopausal bleeding: Society of Radiologists in Ultrasound-Sponsored Consensus Conference statement. *J Ultrasound Med*. 2001 Oct;20(10):1025-36.

Goldstein SR. The role of transvaginal ultrasound or endometrial biopsy in the evaluation of the menopausal endometrium. *Am J Obstet Gynecol*. 2009 Jul;201(1):5-11.

Gull B, Carlsson S, Karlsson B, Ylostalo P, Milsom I, Granberg S. Transvaginal ultrasonography of the endometrium in women with postmenopausal bleeding: is it always necessary to perform an endometrial biopsy? *Am J Obstet Gynecol*. 2000 Mar;182(3):509-15.

Gupta JK, Chien PF, Voit D, Clark TJ, Khan KS. Ultrasonographic endometrial thickness for diagnosing endometrial pathology in women with postmenopausal bleeding: a meta-analysis. *Acta Obstet Gynecol Scand*. 2002 Sep;81(9):799-816.

Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA*. 1994 Feb 2;271(5):389-91.

Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA*. 1994 Mar 2;271(9):703-7.

Yaman C, Habelsberger A, Tews G, Polz W, Ebner T. The role of three-dimensional volume measurement in diagnosing endometrial cancer in patients with postmenopausal bleeding. *Gynecol Oncol*. 2008 Sep;110(3):390-5.

Karlsson B, Granberg S, Wikland M, Ylostalo P, Torvid K, Marsal K, et al. Transvaginal ultrasonography of the endometrium in women with postmenopausal bleeding--a Nordic multicenter study. *Am J Obstet Gynecol*. 1995 May;172(5):1488-94.

Kundel HL, Polansky M. Measurement of observer agreement. *Radiology*. 2003 Aug;228(2):303-8.

La Torre R, De Felice C, De Angelis C, Coacci F, Mastrone M, Cosmi EV. Transvaginal sonographic evaluation of endometrial polyps: a comparison with two dimensional and three dimensional contrast sonography. *Clin Exp Obstet Gynecol*. 1999;26(3-4):171-3.

Laifer-Narin SL, Ragavendra N, Lu DS, Sayre J, Perrella RR, Grant EG. Transvaginal saline hysterosonography: characteristics distinguishing malignant and various benign conditions. *AJR Am J Roentgenol*. 1999 Jun;172(6):1513-20.

Leone FP, Timmerman D, Bourne T, Valentin L, Epstein E, Goldstein SR, et al. 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 Jan;35(1):103-12.

Marret H, Sauget S, Giraudeau B, Body G, Tranquart F. Power Doppler vascularity index for predicting malignancy of adnexal masses. *Ultrasound Obstet Gynecol*. 2005 May;25(5):508-13.

Martinoli C, Derchi LE, Rizzatto G, Solbiati L. Power Doppler sonography: general principles, clinical applications, and future prospects. *Eur Radiol*. 1998;8(7):1224-35.

Merz E, Benoit B, Blaas HG, Baba K, Kratochwil A, Nelson T, et al. Standardization of three-dimensional images in obstetrics and gynecology: consensus statement. *Ultrasound Obstet Gynecol*. 2007 Jun;29(6):697-703.

Odeh M, Vainerovsky I, Grinin V, Kais M, Ophir E, Bornstein J. Three-dimensional endometrial volume and 3-dimensional power Doppler analysis in predicting endometrial carcinoma and hyperplasia. *Gynecol Oncol*. 2007 Aug;106(2):348-53.

Opmeer BC, van Doorn HC, Heintz AP, Burger CW, Bossuyt PM, Mol BW. Improving the existing diagnostic strategy by accounting for characteristics of the women in the diagnostic work up for postmenopausal bleeding. *BJOG*. 2007 Jan;114(1):51-8.

Pairleitner H, Steiner H, Hasenoehrl G, Staudach A. Three-dimensional power Doppler sonography: imaging and quantifying blood flow and vascularization. *Ultrasound Obstet Gynecol*. 1999 Aug;14(2):139-43.

Parsons AK. Evaluation of postmenopausal endometrium. *Ultrasound Obstet Gynecol*. 1998 Nov;12(5):295-300.

Parsons AK, Lense JJ. Sonohysterography for endometrial abnormalities: preliminary results. *J Clin Ultrasound*. 1993 Feb;21(2):87-95.

Pecorelli S. Revised FIGO staging for carcinoma of the vulva, cervix, and endometrium. *Int J Gynaecol Obstet*. 2009 May;105(2):103-4.

Randelzhofer B, Prompeler HJ, Sauerbrei W, Madjar H, Emons G. Value of sonomorphological criteria of the endometrium in women with postmenopausal bleeding: a multivariate analysis. *Ultrasound Obstet Gynecol*. 2002 Jan;19(1):62-8.

Rubin JM, Adler RS. Power Doppler expands standard color capability. *Diagn Imaging (San Franc)*. 1993 Dec;15(12):66-9.

Saha TK, Amer SA, Biss J, Thakare H, Williams S, Farrell CT, et al. The validity of transvaginal ultrasound measurement of endometrial thickness: a comparison of ultrasound measurement with direct anatomical measurement. *BJOG*. 2004 Dec;111(12):1419-24.

Shepherd JH. Revised FIGO staging for gynaecological cancer. *Br J Obstet Gynaecol*. 1989 Aug;96(8):889-92.

Sheth S, Hamper UM, Kurman RJ. Thickened endometrium in the postmenopausal woman: sonographic-pathologic correlation. *Radiology*. 1993 Apr;187(1):135-9.

Smith-Bindman R, Kerlikowske K, Feldstein VA, Subak L, Scheidler J, Segal M, et al. Endovaginal ultrasound to exclude endometrial cancer and other endometrial abnormalities. *JAMA*. 1998 Nov 4;280(17):1510-7.

Takac I. Saline infusion sonohysterography and the risk of malignant extrauterine spread in endometrial cancer. *Ultrasound Med Biol*. 2008 Jan;34(1):7-11.

Van den Bosch T, Verguts J, Daemen A, Gevaert O, Domali E, Claerhout F, et al. Pain experienced during transvaginal ultrasound, saline contrast sonohysterography, hysteroscopy and office sampling: a comparative study. *Ultrasound Obstet Gynecol*. 2008 Mar;31(3):346-51.

Weigel M, Friese K, Strittmatter HJ, Melchert F. Measuring the thickness--is that all we have to do for sonographic assessment of endometrium in postmenopausal women? *Ultrasound Obstet Gynecol*. 1995 Aug;6(2):97-102.

Weber G, Merz E, Bahlmann F, Rosch B. Evaluation of different transvaginal sonographic diagnostic parameters in women with postmenopausal bleeding. *Ultrasound Obstet Gynecol*. 1998 Oct;12(4):265-70.

Widrich T, Bradley LD, Mitchinson AR, Collins RL. Comparison of saline infusion sonography with office hysteroscopy for the evaluation of the endometrium. *Am J Obstet Gynecol.* 1996 Apr;174(4):1327-34.

Ultrasound assessment of endometrial morphology and vascularity to predict endometrial malignancy in women with postmenopausal bleeding and sonographic endometrial thickness ≥ 4.5 mm

G. OPOLSKIENE, P. SLADKEVICIUS and L. VALENTIN

Department of Obstetrics and Gynecology, Malmö University Hospital, Lund University, Malmö, Sweden

ABSTRACT

Objectives *To determine which endometrial morphology characteristics as assessed by gray-scale ultrasound and which endometrial vessel characteristics as assessed by power Doppler ultrasound are useful for discriminating between benign and malignant endometrium in women with postmenopausal bleeding (PMB) and sonographic endometrial thickness ≥ 4.5 mm and to develop logistic regression models to calculate the individual risk of endometrial malignancy in women with PMB, endometrial thickness ≥ 4.5 mm, good visibility of the endometrium and detectable Doppler signals in the endometrium.*

Methods *Of 223 consecutive patients with PMB and sonographic endometrial thickness ≥ 4.5 mm, 120 fulfilled our inclusion criteria. They underwent transvaginal gray-scale and power Doppler ultrasound examination, which was videotaped for later analysis by two examiners with more than 15 years' experience in gynecological ultrasonography. They independently assessed endometrial morphology and vascularity using predetermined criteria. Their agreed-upon description was compared with the histological diagnosis. Univariate and multivariate logistic regression analyses were used. The best diagnostic test was defined as the one with the largest area under the receiver–operating characteristics curve (AUC).*

Results *Thirty (25%) endometria were malignant. Interobserver agreement for the description of endometrial morphology and vascularity was moderate to good (Kappa 0.49–0.78). The best ultrasound variables to predict malignancy were heterogeneous endometrial echogenicity (AUC 0.83), endometrial thickness (AUC 0.80), and irregular branching of endometrial blood vessels (AUC 0.77). A logistic regression model including endometrial thickness and heterogeneous endometrial echogenicity had an AUC of 0.91. Its mathematically best risk cut-off yielded a positive likelihood ratio of 4.4, and a negative likelihood ratio of 0.1. Adding Doppler information to the model improved diagnostic performance marginally (AUC 0.92).*

Conclusions *In selected high-risk women with PMB and an endometrial thickness of ≥ 4.5 mm, calculation of the individual risk of endometrial malignancy using regression models including gray-scale and Doppler characteristics can be used to tailor management. These models would need to be tested prospectively before introduction into clinical practice. Copyright © 2007 ISUOG. Published by John Wiley & Sons, Ltd.*

INTRODUCTION

There is strong scientific evidence that endometrial thickness as measured by ultrasonography can discriminate between women with postmenopausal bleeding at low and high risk of endometrial cancer, women with endometrial thickness ≤ 4 mm having a low risk and those with endometrial thickness ≥ 5 mm having a high risk[1]. While it seems to be safe to refrain from endometrial sampling in women with endometrial thickness ≤ 4 mm, women with endometrial thickness ≥ 5 mm should have their endometrium sampled[2]. Even within the group of women at high risk with respect to endometrial thickness, it would be of clinical value to be able to identify

ORIGINAL PAPER

those at the lowest risk and those at the highest risk for endometrial cancer, because this would make it possible to individualize management. We wanted to explore whether variables other than endometrial thickness, e.g. the gray-scale ultrasound morphology of the endometrium, the vascularization of the endometrium as assessed by Doppler ultrasound, or clinical variables could help in discriminating between benign and malignant endometrium in women with postmenopausal bleeding and a thick endometrium.

The aims of this study were (1) to determine which endometrial morphology characteristics as assessed by gray-scale ultrasonography and which endometrial vessel characteristics as assessed by power Doppler ultrasound are useful for discriminating between benign and malignant endometrium in women with postmenopausal bleeding and sonographic endometrial thickness ≥ 4.5 mm and (2) to develop logistic regression models to calculate the individual risk of endometrial malignancy in women with postmenopausal bleeding, endometrial thickness ≥ 4.5 mm, good visibility of the endometrium and detectable Doppler signals in the endometrium.

PATIENTS AND METHODS

Consecutive patients with postmenopausal bleeding, endometrial thickness ≥ 4.5 mm at transvaginal ultrasound examination (measurement taken using the double layer technique[3]) and without fluid in the intrauterine cavity underwent extended ultrasound examination as described below. A woman was considered to be postmenopausal if she reported absence of menstruation for at least 1 year after the age of 40 provided that the amenorrhea was not explained by medication or disease. Postmenopausal bleeding was defined as any vaginal bleeding in a postmenopausal woman not on hormone replacement therapy (HRT), or unscheduled vaginal bleeding in a postmenopausal woman on HRT. Transvaginal sonography was carried out by one of two examiners using a Sequoia 512 ultrasound system (Siemens Medical Solutions Inc., Ultrasound Division, Mountain View, CA, USA) equipped with a 5–8-MHz transvaginal transducer. All the women were examined in the lithotomy position with an empty bladder. First, conventional gray-scale ultrasound examination of the uterus was performed, and then power Doppler ultrasound examination was carried out using predetermined, standardized settings (frequency, 6 MHz; power Doppler gain, 50; dynamic range, 10 dB; edge, 1; persistence, 2; color map, 1; gate, 2; filter, 3). The examinations were videotaped for later analysis.

Approximately 12 months after the collection of data had been completed, two examiners, both of whom had more than 15 years' experience in gynecological scanning, reviewed the videotapes. They assessed endometrial morphology and vascularity using a fixed study protocol. This protocol included predetermined definitions of endometrial morphology characteristics and endometrial blood vessel characteristics. To minimize bias when analyzing the power Doppler images, the gray-scale images were analyzed several months before the power Doppler images. The analysis of gray-scale ultrasound endometrial morphology included visual evaluation of the following:

- presence of bright line(s) separating the endometrial echo from the myometrium (single line, double lines, no lines);
- presence of middle echo (yes, no);
- regularity of the endometrial–myometrial border (regular, irregular, impossible to evaluate);
- internal endometrial structure (hyperechogenic, hypoechogenic, isoechogenic, cystic, impossible to evaluate); and
- homogeneity of endometrial echogenicity (homogeneous, heterogeneous, impossible to evaluate).

Examples of these morphological characteristics are presented in Figure 1.

Analysis of the videotaped power Doppler ultrasound examinations included visual evaluation of the following:

- number of blood vessels crossing the myometrial–endometrial border (one, two or many);
- size of blood vessels (small or large, any large vessel having precedence over small ones);
- regularity of vessel branching (regular, irregular);
- presence of large areas of color, i.e. 'color splashes' (yes, no); and
- presence of area(s) of densely packed blood vessels (yes, no).

In addition, endometrial vascularity was classified as multiple vascular pattern (A), single vascular pattern (B), or scattered vascular pattern (C) as proposed by Alcazar *et al.*[4]. The vascular characteristics are illustrated in Figure 2.

To determine interobserver reproducibility of the evaluation of endometrial morphology and vascularity the two observers performed their evaluations of the videotapes independently of each other. Any disagreement in their results was resolved by discussion between the two observers while re-viewing the tapes together. Their agreed-upon classification was used for statistical analysis and the results of the ultrasound examinations were compared with those of histological examination of the respective surgical specimens obtained by dilatation and curettage, hysteroscopic resection or hysterectomy. Staging of malignant tumors was done by the attending physician in accordance with the classification system recommended by the International Federation of Gynecology and Obstetrics[5]. Exclusion criteria were the absence of power Doppler signals in the endometrium, power Doppler artifacts making the power Doppler image uninterpretable, technical problems, e.g. large myomas, making it impossible to study in detail the gray-scale ultrasound morphology and/or the vascularity of the endometrium, incomplete videotaping, absence of histopathological
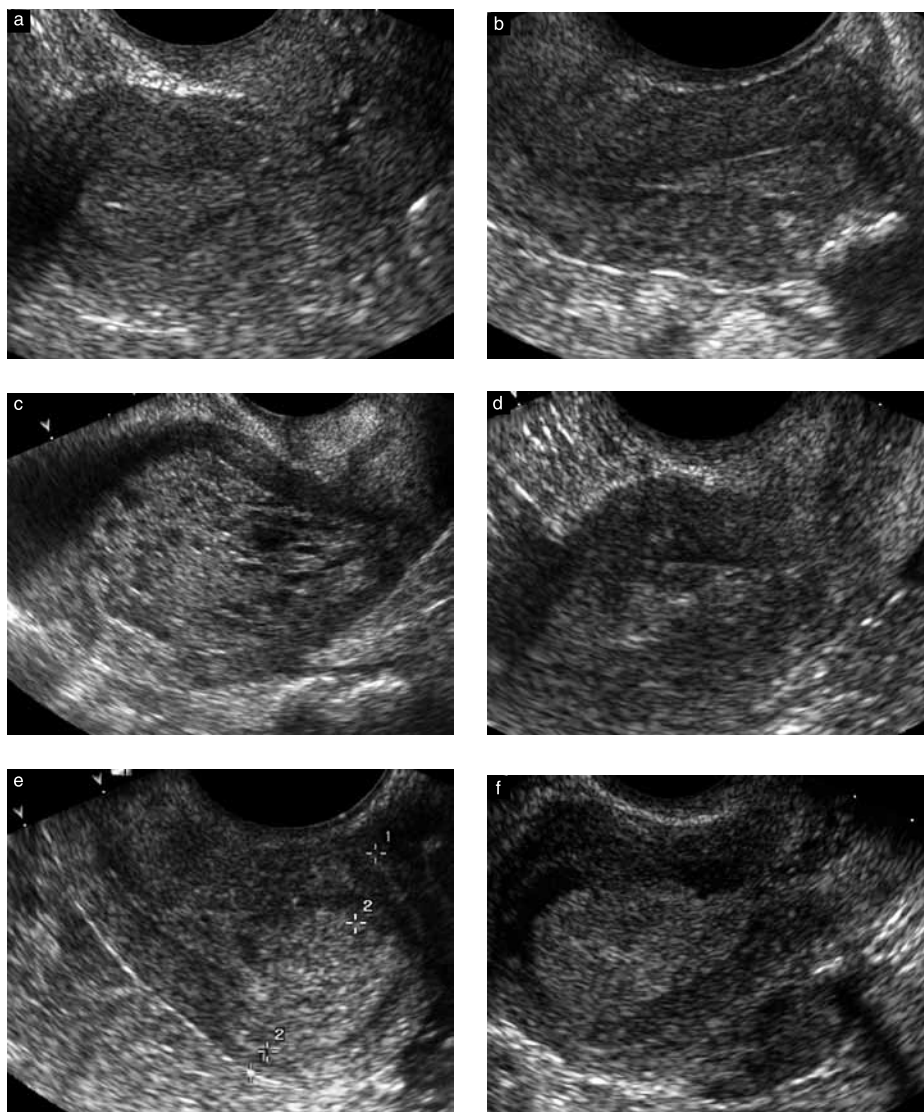
Figure 1 Illustration of gray-scale ultrasound characteristics of the endometrium, showing: (a) homogeneous endometrial echogenicity with middle echo (histopathological diagnosis: benign estrogen-influenced endometrium); (b) lines separating the endometrial echo from the myometrium (histopathological diagnosis: benign polyp); (c) cystic endometrial structure and regular endometrial–myometrial borders with lines separating the endometrial echo from the myometrium (histopathological diagnosis: benign polyp); (d) heterogeneous endometrial echogenicity with irregular endometrial–myometrial border anteriorly (histopathological diagnosis: adenocarcinoma); (e) irregular endometrial–myometrial border mainly posteriorly (histopathological diagnosis: adenocarcinoma), and (f) heterogeneous endometrial echogenicity and irregular endometrial–myometrial borders (histopathological diagnosis: adenocarcinoma).
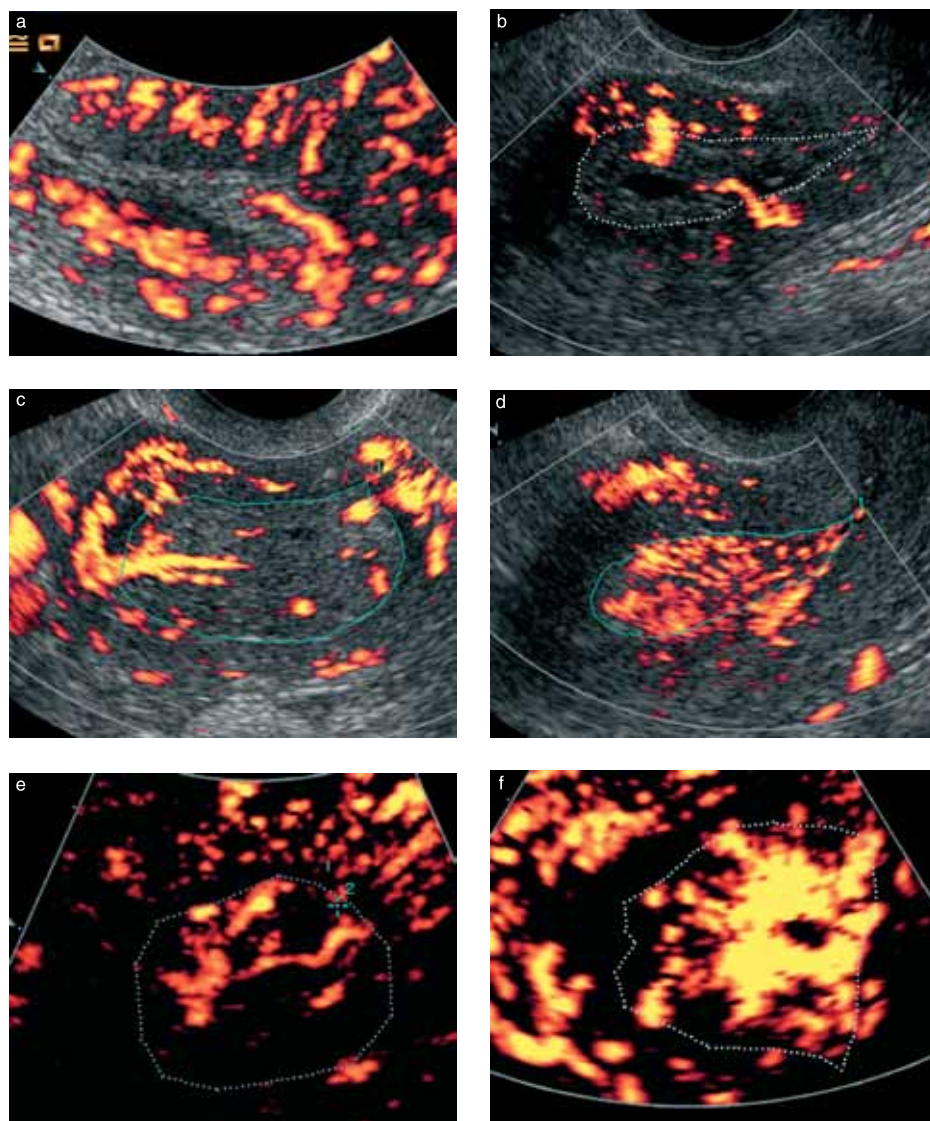
Figure 2 Illustration of vascular characteristics showing: (a) one blood vessel crossing the myometrial–endometrial border (histopathological diagnosis: benign polyp); (b) two regular blood vessels crossing the myometrial–endometrial border of the cystic endometrium (histopathological diagnosis: benign polyp); (c) regular branching of endometrial blood vessels (histopathological diagnosis: benign polyp); (d) multiple, densely packed endometrial blood vessels (histopathological diagnosis: adenocarcinoma); (e) irregular branching of endometrial blood vessels (histopathological diagnosis: adenocarcinoma); and (f) the presence of color 'splashes' (histopathological diagnosis: adenocarcinoma). The dotted line in images b, c, d, e and f outlines the endometrium. The black background in images e and f resulted from elimination of the gray-scale image for better visualization of the vessels.

diagnosis, or histopathological diagnosis obtained only by an outpatient endometrial sampling device (e.g. Pipelle® or Endorette®). The reason for excluding samples obtained only by an outpatient endometrial sampling device was that we wanted to be sure that only representative samples were used to establish the final diagnosis[6,7].

Statistical calculations were undertaken using the Statistical Package for the Social Sciences, version 12.02 (SPSS Inc., Chicago, IL, USA) and the statistical software StatXact, version 4 (Cytel Inc., Cambrige, MA, USA). The Mann–Whitney *U*-test was used to determine the statistical significance of differences in age and endometrial thickness, and Fisher's exact test was used to determine the statistical significance of differences in the use of HRT. Cohen's Kappa coefficient was calculated to assess interobserver agreement, values of 0.81−1 being taken to indicate almost perfect agreement, values of 0.61−0.8 good agreement and values of 0.41−0.6 moderate agreement[8]. The statistical significance of a possible relationship between endometrial malignancy and clinical variables (age and use of HRT) and ultrasound variables was determined using univariate logistic regression analysis with the likelihood ratio test. Two-tailed $P < 0.05$ was considered statistically significant.

Multivariate logistic regression analysis was used to build models to predict malignancy. To avoid overfitting, a maximum of three predicting variables were allowed in a model, the likelihood ratio test yielding $P < 0.05$ being the criterion for including a variable in a model. Building logistic regression models we first determined whether any clinical variable (age, use of HRT), gray-scale variable or power Doppler variable added useful information to endometrial thickness. We then studied the effect of adding power Doppler variables to the best gray-scale models that included endometrial thickness as one of the gray-scale variables. We also built models by adding power Doppler variables to the best gray-scale models that did not include endometrial thickness as a variable.

The application of the regression equations to data from each woman gave the probability for that woman to have an endometrial malignancy, the probability ranging from 0 to 1. Receiver–operating characteristics (ROC) curves were drawn for single predicting variables as well as for regression equations to evaluate their diagnostic ability. The area under the ROC curve and the 95% CI of this area were calculated. If the lower limit of the CI for the area under the ROC curve was > 0.5, the diagnostic test was considered to have discriminatory potential. For continuous variables the ROC curves were also used to determine the mathematically best cut-off value to predict malignancy for each diagnostic test (single variables as well as logistic regression models), the mathematically best cut-off value being defined as that corresponding to the point on the ROC curve situated furthest away from the reference line. The sensitivity, specificity, positive and negative likelihood ratios (LR) of the mathematically best cut-off value with their 95% CIs were then calculated. We

defined the best diagnostic test as the one with the largest area under the ROC curve.

The Ethics Committee of Lund University approved the study protocol and informed consent was obtained from all the participants after the nature of the procedures had been fully explained.

RESULTS

A total of 223 consecutive women with postmenopausal bleeding and endometrial thickness ≥ 4.5 mm were examined in our ultrasound unit. Of these, 103 women were excluded from this study for the following reasons: free fluid in the endometrial cavity ($n = 12$), absence of power Doppler signals in the endometrium ($n = 14$), large myomas obscuring the view of the endometrium ($n = 9$), absence of histological diagnosis ($n = 13$), histological diagnosis evaluated only from a specimen taken by a simple outpatient sampling device ($n = 13$) and examinations not properly videotaped ($n = 42$). Among the women excluded, 49 (48%) had benign endometrium and six (6%) had malignant endometrium, 22 (21%) underwent endometrial sampling only using a simple outpatient endometrial sampling device (11 of these had benign endometrium, while in the remaining 11 the endometrial samples were insufficient for diagnosis), three (3%) underwent dilatation and curettage but with insufficient material for diagnosis, and 23 (22%) did not undergo any endometrial sampling. Median endometrial thickness in the women excluded was 8.7 (range, 4.5–38.6) mm.

Of the 120 women included, 90 had benign and 30 had malignant endometrium. Histological diagnoses are shown in Table 1. Twenty patients (67%) had Stage I, four (13%) had Stage II, four (13%) had Stage III, one (3%) had Stage IV endometrial cancer, and one woman did not undergo a proper staging procedure because of high operative risk. Women with malignant endometrium were older than those with benign endometrium (median 73 (range, 56–85) years vs. median 63 (range, 43–90) years; $P = 0.023$) and they had thicker endometrium (median 17.6 (range 6.7–50.0) mm, vs. 10.2 (range, 4.6–30.1) mm; $P = 0.0005$). Seventeen women (14%) were on continuous combined or sequential HRT, 17

Table 1 Histopathological diagnoses

Finding	n (%)
Benign ($n = 90$)	
Polyp	62 (52)
Atrophic endometrium	11 (9)
Estrogen-influenced endometrium	10 (8)
Hyperplasia	
Without atypia	5 (4)
With atypia	2 (2)
Malignant ($n = 30$)	
Adenocarcinoma	27 (22.5)
Carcinosarcoma	3 (2.5)
Total	120 (100)

(14%) were using low-dose oral estrogens or local estrogens, 84 (70%) used no hormonal therapy at all, and for two women (2%) information on the use of HRT was lacking. The proportion of women using HRT did not differ significantly between women with benign and malignant endometrium (16% (14/88) vs. 10% (3/30); $P = 0.56$).

The sensitivity, specificity, positive and negative LR, and area under the ROC curve for age, HRT and ultrasound variables are shown in Table 2. The gray-scale ultrasound morphology variable that best predicted malignancy was heterogeneous endometrial echogenicity (area under the ROC curve 0.83, sensitivity 73%, specificity 92%, positive LR 9.4, negative LR 0.3), and the power Doppler ultrasound variable that best predicted malignancy was irregular branching of endometrial blood vessels (area under the ROC curve 0.77, sensitivity 60%, specificity 94%, positive LR 10.8, negative LR 0.4).

The performance of the logistic regression models and the mathematical formulae of the best models are shown in Table 3. None of the clinical variables entered a logistic regression model to predict malignancy. The best logistic regression model with only two ultrasound variables was a model including endometrial thickness and heterogeneous echogenicity of the endometrium (area under the ROC curve 0.91). The diagnostic performance improved marginally when we added Doppler information (areas of densely packed blood vessels or irregular branching of endometrial blood vessels) to this model (area under the ROC curve 0.92).

Interobserver agreement for evaluation of endometrial vascularity (Cohen's kappa 0.49–0.78) was superior to that of evaluation of gray-scale endometrial ultrasound morphology (Cohen's kappa 0.50–0.66; Table 4). Agreement was best for color 'splashes' in the endometrium (Kappa index 0.78), areas of densely packed blood vessels in the endometrium (Kappa index 0.75), branching of endometrial blood vessels (Kappa index 0.67), and homogeneity of endometrial echogenicity (Kappa index 0.66).

DISCUSSION

The results of our study show that in a selected group of women with postmenopausal bleeding, endometrial thickness ≥ 4.5 mm, good visibility of the endometrium and detectable Doppler signals in the endometrium, endometrial morphology as assessed by gray-scale ultrasonography and endometrial vessel characteristics as assessed by power Doppler ultrasonography are independently related to endometrial malignancy. Moreover, both add useful information to sonographic endometrial thickness as an indicator of the risk for endometrial carcinoma. The reason why women without detectable Doppler signals in the endometrium were excluded from our study is that vessel characterization is not possible in endometria without detectable Doppler signals.

The single best ultrasound variable for predicting endometrial malignancy was heterogeneous endometrial echogenicity. It was superior to endometrial thickness, but it changed the odds of malignancy only moderately (positive LR 9.4, negative LR 0.3[9,10]). The internal endometrial structure most suggestive of malignancy was subjectively perceived as being 'moth eaten' (Figures 1 d and f). Others have also reported heterogeneous

Table 2 Sensitivity and specificity with regard to malignancy, positive and negative likelihood ratios, and area under the receiver–operating characteristics (ROC) curve of clinical and ultrasound variables

Variable	ROC curve area estimate	Optimal probability cut-off*	Sensitivity (% (n))	Specificity (% (n))	LR+	LR−	P†
Endometrial thickness	0.80	15 mm	73 (22/30)	77 (69/90)	3.1	0.3	0.0005
Age	0.64	61 years	87 (26/30)	40 (36/90)	1.4	0.3	0.041
Hormone replacement therapy	0.47	—	10 (3/30)	84 (74/88)	0.6	1.1	0.41
Gray-scale analysis							
Heterogeneous echogenicity	0.83	—	73 (22/30)	92 (83/90)	9.4	0.3	0.0005
Irregular border	0.71	—	67 (20/30)	76 (68/90)	2.7	0.4	0.0005
Cystic endometrium	0.67	—	23 (7/30)	42 (38/90)	0.4	1.8	0.001
Hyperechogenic endometrium	0.67	—	53 (16/30)	62 (56/90)	1.4	0.8	0.137
Hypoechogenic endometrium	0.54	—	10 (3/30)	99 (89/90)	9.0	0.9	0.032
Isoechogenic endometrium	0.54	—	13 (4/30)	96 (86/90)	3.0	0.9	0.114
No lines	0.50	—	50 (15/30)	50 (45/90)	1.0	1.0	1.0
Power Doppler analysis							
Irregular branching	0.77	—	60 (18/30)	94 (85/90)	10.8	0.4	0.0005
Areas of densely packed vessels	0.76	—	67 (20/30)	84 (76/90)	4.3	0.4	0.0005
Color 'splashes'	0.72	—	60 (18/30)	84 (76/90)	3.9	0.5	0.0005
Many vessels	0.68	—	83 (25/30)	52 (47/90)	1.7	0.3	0.0005
Vascular pattern A‡	0.67	—	80 (24/30)	54 (49/90)	1.8	0.4	0.001
Branching of vessels	0.67	—	87 (26/30)	47 (42/90)	1.6	0.3	0.001
Large vessels	0.57	—	90 (27/30)	23 (21/90)	1.2	0.4	0.094

A more exhaustive table including the 95% confidence intervals of all measures of test performance presented here can be found in the electronic version of the paper (see Table S1). *Values above the cut-off indicate malignancy. †Univariate logistic regression with likelihood ratio test. ‡Vascular pattern A according to Alcazar *et al.*[4]. LR, likelihood ratio.

Table 3 Multivariate logistic regression models to predict malignancy

Model	ROC curve area estimate	Optimal probability cut-off*	Sensitivity (% (n))	Specificity (% (n))	LR+	LR−
Endometrial thickness and:						
Heterogeneous echogenicity[1]	0.91	0.12	93 (28/30)	79 (71/90)	4.4	0.1
Irregular border	0.84	0.31	80 (24/30)	87 (78/90)	6.0	0.2
Cystic endometrium	0.84	0.27	77 (23/30)	86 (77/90)	5.3	0.3
Hyperechogenic endometrium	0.81	0.31	70 (21/30)	87 (78/90)	5.3	0.3
Irregular branching[2]	0.86	0.29	77 (23/30)	88 (79/90)	6.3	0.3
Areas of densely packed vessels[3]	0.85	0.28	83 (25/30)	82 (74/90)	4.7	0.2
Color 'splashes'	0.83	0.33	77 (23/30)	82 (74/90)	4.3	0.3
Many vessels	0.82	0.24	80 (24/30)	77 (69/90)	3.4	0.3
Pattern A†	0.81	0.26	77 (23/30)	77 (69/90)	3.3	0.3
Endometrial thickness, heterogeneous echogenicity and:						
Areas of densely packed vessels[4]	0.92	0.28	83 (25/30)	89 (80/90)	7.5	0.2
Irregular branching	0.92	0.15	87 (26/30)	83 (75/90)	5.2	0.2
Color 'splashes'	0.91	0.31	80 (24/30)	90 (81/90)	8.0	0.2
Endometrial thickness, irregular border and:						
Areas of densely packed vessels	0.86	0.44	73 (22/30)	92 (83/90)	9.4	0.3
Color 'splashes'	0.85	0.52	67 (20/30)	96 (86/90)	15.0	0.3
Heterogeneous echogenicity and:						
Areas of densely packed vessels[5]	0.89	0.41	73 (22/30)	92 (83/90)	9.4	0.3
Color 'splashes'	0.87	0.42	73 (22/30)	92 (83/90)	9.4	0.3
Irregular branching	0.87	0.23	80 (24/30)	88 (79/90)	6.5	0.2
Branching	0.86	0.32	73 (22/30)	92 (83/90)	9.4	0.3
Pattern A†	0.86	0.34	73 (22/30)	92 (83/90)	9.4	0.3
Irregular endometrial–myometrial border and:						
Areas of densely packed vessels	0.80	0.56	57 (17/30)	97 (87/90)	17.0	0.4
Color 'splashes'	0.78	0.55	50 (15/30)	98 (88/90)	22.5	0.5
Branching	0.77	0.40	63 (19/30)	88 (79/90)	5.2	0.4
Many vessels	0.75	0.39	67 (20/30)	84 (76/90)	4.3	0.4
Pattern A†	0.75	0.40	67 (20/30)	87 (78/90)	5.0	0.4

A more exhaustive table including the 95% confidence intervals of all measures of test performance presented here can be found in the electronic version of the paper (see Table S2). *Values above the probability cut-off indicate malignancy. †Vascular pattern A according to Alcazar *et al.*[4]. LR, likelihood ratio; ROC, receiver–operating characteristics.

The probability of malignancy is calculated as $[e^z/(1 + e^z)]$ where $e = 2.718$ (base value of natural logarithms) and z is calculated for each logistic regression model as follows:

[1] $z = -3.988 + (0.125 \times$ endometrial thickness in mm$) + (3.012 \times$ echogenicity coded 0 if homogeneous and 1 if heterogeneous)

[2] $z = -4.151 + (0.156 \times$ endometrial thickness in mm$) + (2.886 \times$ branching coded 0 if regular and 1 if irregular)

[3] $z = -4.411 + (0.167 \times$ endometrial thickness in mm$) + (2.151 \times$ areas of densely packed vessels coded 0 if absent and 1 if present)

[4] $z = -4.462 + (0.113 \times$ endometrial thickness in mm$) + (2.832 \times$ echogenicity coded 0 if homogeneous and 1 if heterogeneous$) + (1.907 \times$ areas of densely packed vessels coded 0 if absent and 1 if present)

[5] $z = -3.037 + (3.243 \times$ echogenicity coded 0 if homogeneous and 1 if heterogeneous$) + (2.050 \times$ areas of densely packed vessels coded 0 if absent and 1 if present).

endometrial structure to be associated with endometrial malignancy[11]. Irregular endometrial–myometrial border was also a sign of endometrial cancer, but in agreement with others we found it to be a poorer predictor of malignancy than heterogeneous endometrial structure[12]. The presence of irregular branching of endometrial blood vessels increased the odds of malignancy almost 11-fold. Epstein and Valentin also noted that irregular vessel branching was more common in malignant than in benign endometria[11]. The presence of areas with densely packed blood vessels and 'color splashes' within the endometrium or in the endometrial–myometrial border increased the odds of malignancy 4-fold. Densely packed vessels or color splashes may not necessarily reflect microvessel density in the endometrium, but it is nonetheless interesting that endometrial carcinoma is associated with increased microvessel counts[13]. The vascular patterns A, B and C, which worked well in the study of Alcazar *et al.*[4], did not perform well as predictors of endometrial malignancy in our hands. The discrepancy may be explained by fundamental differences in study design.

The crucial question is whether or not gray-scale ultrasound morphology and endometrial vascularity as assessed by Doppler ultrasound are superior to, or add to, simple sonographic endometrial thickness measurements in the prediction of endometrial cancer. Even in our selected high-risk group of patients with postmenopausal bleeding and endometrial thickness ≥ 4.5 mm, the risk of malignancy increased with increasing endometrial thickness, and endometrial thickness was a fairly good predictor of malignancy. Heterogeneous endometrial echogenicity was only slightly superior to endometrial

Table 4 Interobserver agreement

Parameter	Cohen's kappa	Agreement (%)
Gray-scale analysis		
Homogeneity of endometrial echogenicity	0.66	87
Endometrial echogenicity (cystic, hyper-, hypo- or isoechogenic)	0.61	77
Endometrial lines	0.56	73
Regularity of endometrial–myometrial border	0.50	76
Power Doppler analysis		
Color 'splashes'	0.78	92
Areas of densely packed vessels	0.75	90
Branching of vessels	0.67	84
Branching regularity	0.59	74
Size of vessels	0.52	85
Vascular pattern A, B, C*	0.49	69
Number of vessels	0.49	68

*Vascular pattern A, B, C according to Alcazar *et al.*[4].

thickness, but it did add information to endometrial thickness in a logistic regression model. A model including endometrial thickness and heterogeneous endometrial echogenicity was the best one for predicting malignancy. Using the mathematically optimal risk cut-off of this model misclassified only two of the malignant endometria and only nine of the benign ones. Adding Doppler variables to this model improved the overall diagnostic performance only marginally, and the use of the respective optimal risk cut-offs of the models including Doppler variables did not result in more endometria being correctly classified. Models not containing endometrial thickness but only heterogeneous endometrial echogenicity and Doppler variables also performed well.

A problem with the evaluation of ultrasound images is its subjectivity. Indeed interobserver agreement for categorizing gray-scale ultrasound and Doppler findings was only moderate or good. Some Doppler variables were more reproducible than even the most reproducible gray-scale ultrasound variable. In a reproducibility study by Alcazar *et al.* interobserver agreement for evaluating the vessel pattern in the endometrium was good between two experts but not between less experienced examiners[14].

Quite clearly, ultrasound evaluation of the endometrium in women with postmenopausal bleeding starts with a proper measurement of the endometrial thickness. In women with postmenopausal bleeding endometrial thickness ≤ 4 mm decreases the odds of malignancy 10-fold, the risk of endometrial cancer in such women varying between $1:1000$ and $1:100$[1]. It is important to bear in mind that our study group included only women with postmenopausal bleeding at high risk of endometrial cancer, i.e. those with endometrial thickness ≥ 4.5 mm, and among these only those without fluid in the uterine cavity, well visible endometrium and detectable power Doppler signals in the endometrium without power Doppler artifacts. Our results are only applicable to similar populations. The rationale for

studying only a high-risk group with endometrial thickness ≥ 4.5 mm is that first, it would be very difficult to assess endometrial gray-scale and vessel morphology in an endometrium ≤ 4.4 mm, and second further risk assessment in women with postmenopausal bleeding and endometrial thickness ≤ 4.4 mm seems unnecessary, because in these women the risk of endometrial cancer is so low that it is safe to refrain from endometrial sampling[2]. In the high-risk group with endometrial thickness ≥ 4.5 mm, however, a differentiation of risk would allow individualized management. For example, in a woman with an estimated risk of endometrial cancer $< 1:100$ (calculated using our best logistic regression model) at high operative risk it might be appropriate to refrain from endometrial sampling, at least if cervical stenosis – or other factors – makes it impossible to obtain an endometrial sample using an outpatient sampling device. On the other hand, a high risk of malignancy would support not delaying a reliable diagnostic procedure. Needless to say, our logistic regression models need to be tested prospectively before they can be used in clinical practice.

REFERENCES

1. Smith-Bindman R, Kerlikowske K, Feldstein VA, Subak L, Scheidler J, Segal M, Brand R, Grady D. Endovaginal ultrasound to exclude endometrial cancer and other endometrial abnormalities. *JAMA* 1998; **280**: 1510–1517.
2. Gull B, Karlsson B, Milsom I, Granberg S. Can ultrasound replace dilation and curettage? A longitudinal evaluation of postmenopausal bleeding and transvaginal sonographic measurement of the endometrium as predictors of endometrial cancer. *Am J Obstet Gynecol* 2003; **188**: 401–408.
3. Epstein E, Valentin L. Managing women with post-menopausal bleeding. *Best Pract Res Clin Obstet Gynaecol* 2004; **18**: 125–143.
4. Alcazar JL, Castillo G, Minguez JA, Galan MJ. Endometrial blood flow mapping using transvaginal power Doppler sonography in women with postmenopausal bleeding and thickened endometrium. *Ultrasound Obstet Gynecol* 2003; **21**: 583–588.
5. Shepherd JH. Revised FIGO staging for gynaecological cancer. *Br J Obstet Gynaecol* 1989; **96**: 889–892.
6. Epstein E, Skoog L, Milsom L. Comparison of Endorette and dilatation and curettage for sampling of the endometrium in women with postmenopausal bleeding. *Acta Obstet Gynecol Scand* 2001; **80**: 959–964.
7. Farrell T, Jones N, Owen P, Baird A. The significance of an 'insufficient' Pipelle sample in the investigation of postmenopausal bleeding. *Acta Obstet Gynecol Scand* 1999; **78**: 810–812.
8. Kundel HL, Polansky M. Measurement of observer agreement. *Radiology* 2003; **228**: 303–308.

9. Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA* 1994; **271**: 389–391.

10. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA* 1994; **271**: 703–707.

11. Epstein E, Valentin L. Gray-scale ultrasound morphology in the presence or absence of intrauterine fluid and vascularity as assessed by color Doppler for discrimination between benign and malignant endometrium in women with postmenopausal bleeding. *Ultrasound Obstet Gynecol* 2006; **28**: 89–95.

12. Randelzhofer B, Prompeler HJ, Sauerbrei W, Madjar H, Emons G. Value of sonomorphological criteria of the endometrium in women with postmenopausal bleeding: a multivariate analysis. *Ultrasound Obstet Gynecol* 2002; **19**: 62–68.

13. Abulafia O, Sherer DM. Angiogenesis of the endometrium. *Obstet Gynecol* 1999; **94**: 148–153.

14. Alcazar JL, Ajossa S, Floris S, Bargellini R, Gerada M, Guerriero S. Reproducibility of endometrial vascular patterns in endometrial disease as assessed by transvaginal power Doppler sonography in women with postmenopausal bleeding. *J Ultrasound Med* 2006; **25**: 159–163.

SUPPLEMENTARY MATERIAL ON THE INTERNET

The following material is available from the Journal homepage: http://www.interscience.wiley.com/jpages/0960-7692/suppmat (restricted access)

Table S1 Sensitivity and specificity with regard to malignancy, positive and negative likelihood ratios, and area under the receiver–operating characteristics curve of clinical and ultrasound variables. 95% CIs are shown in addition to the data provided in Table 2 of the print version.

Table S2 Multivariate logistic regression models to predict malignancy. 95% CIs are shown in addition to the data provided in Table 3 of the print version.

Table S1 Sensitivity and specificity with regard to malignancy, positive and negative likelihood ratios, and area under the receiver operating characteristics curve of clinical and ultrasound variables. 95% CIs are shown in addition to the data provided in the Table 2.

Variable	ROC area Estimate (95% CI)	Optimal cut-off*	Sensitivity %, (n) (95% CI)	Specificity %, (n) (95% CI)	LR+ (95% CI)	LR- (95% CI)	P-value**
Endometrial thickness	0.80 (0.71 – 0.89)	15 mm	73 (22/30) (54 – 88)	77 (69/90) (67 – 85)	3.1 (1.8 – 5.9)	0.3 (0.1 – 0.6)	0.0005
Age	0.64 (0.54 – 0.74)	61 years	87 (26/30) (69 – 96)	40 (36/90) (30 – 51)	1.4 (1.0 – 2.0)	0.3 (0.03 – 0.9)	0.041
Hormone replacement therapy	0.47 (0.35 – 0.59)	-	10 (3/30) (2 – 27)	84 (74/88) (75 – 91)	0.6 (0.03 – 2.7)	1.1 (0.8 – 1.3)	0.41
Gray scale analysis							
Heterogeneous echogenicity	0.83 (0.73 – 0.93)	-	73 (22/30) (54 – 88)	92 (83/90) (85 – 97)	9.4 (3.9 – 33.1)	0.3 (0.1 – 0.5)	0.0005
Irregular border	0.71 (0.60 – 0.82)	-	67 (20/30) (47 – 83)	76 (68/90) (65 – 84)	2.7 (1.5 – 5.2)	0.4 (0.2 – 0.8)	0.0005
Cystic endometrium	0.67 (0.56 – 0.78)	-	23 (7/30) (10 – 42)	42 (38/90) (32 – 53)	0.4 (0.1 – 0.8)	1.8 (1.2 – 2.9)	0.001
Hyperechogenic endometrium	0.67 (0.56 – 0.77)	-	53 (16/30) (34 – 72)	62 (56/90) (51 – 72)	1.4 (0.7 – 2.6)	0.8 (0.4 – 1.2)	0.137
Hypoechogenic endometrium	0.54 (0.42 – 0.67)	-	10 (3/30) (2 – 27)	99 (89/90) (94 – 100)	9.0 (0.3–31250§)	0.9 (0.7 – 1.0)	0.032
Isoechogenic endometrium	0.54 (0.42 – 0.67)	-	13 (4/30) (4 – 31)	96 (86/90) (89 – 99)	3.0 (0.2 – 38.3)	0.9 (0.7 – 1.1)	0.114
No lines	0.50 (0.38 – 0.62)	-	50 (15/30) (31 – 69)	50 (45/90) (39 – 61)	1.0 (0.5 – 1.7)	1.0 (0.5 – 1.7)	1.0
Power Doppler analysis							
Irregular branching	0.77 (0.66 – 0.89)	-	60 (18/30) (41 – 77)	94 (85/90) (88 – 98)	10.8 (3.7 – 58.1)	0.4 (0.2 – 0.6)	0.0005
Areas of densely packed vessels	0.76 (0.65 – 0.87)	-	67 (20/30) (47 – 83)	84 (76/90) (75 – 91)	4.3 (2.1 – 10.0)	0.4 (0.2 – 0.7)	0.0005

Continued

Table S1 continued

Variable	ROC area Estimate (95% CI)	Optimal cut-off*	Sensitivity %, (n) (95% CI)	Specificity %, (n) (95% CI)	LR+ (95% CI)	LR- (95% CI)	P-value**
Color 'splashes'	0.72 (0.61 − 0.84)	-	60 (18/30) (41 − 77)	84 (76/90) (75 − 91)	3.9 (1.8 − 9.3)	0.5 (0.2 − 0.7)	0.0005
Many vessels	0.68 (0.57 − 0.78)	-	83 (25/30) (65 − 94)	52 (47/90) (41 − 63)	1.7 (1.2 − 2.6)	0.3 (0.1 − 0.7)	0.0005
Vascular pattern A***	0.67 (0.56 − 0.78)	-	80 (24/30) (61 − 92)	54 (49/90) (44 − 65)	1.8 (1.2 − 2.7)	0.4 (0.1 − 0.8)	0.001
Branching of vessels	0.67 (0.56 − 0.77)	-	87 (26/30) (69 − 96)	47 (42/90) (36 − 57)	1.6 (1.1 − 2.3)	0.3 (0.03 − 0.8)	0.001
Large vessels	0.57 (0.45 − 0.68)	-	90 (27/30) (73 − 98)	23 (21/90) (15 − 33)	1.2 (0.9 − 1.5)	0.4 (0.02 − 1.6)	0.094

LR, likelihood ratio; ROC, receiver operating characteristic; CI, confidence interval;
*Values above the cut-off indicate malignancy; **Univariate logistic regression with likelihood ratio test; ***Vascular pattern A according to Alcazar et al[4], §Unstable LR+ because of high specificity (89/90)

69

Table S2 Multivariate logistic regression models to predict malignancy. 95% CIs are shown in addition to the data provided in the Table 3.

Model	ROC area Estimate (95% CI)	Optimal probability cutoff	Sensitivity, %, (n) (95% CI)	Specificity %, (n) (95% CI)	LR+ (95% CI)	LR- (95% CI)
Endometrial thickness and						
heterogenous echogenicity[1]	0.91 (0.84 – 0.97)	0.12	93 (28/30) (78 – 99)	79 (71/90) (69 – 87)	4.4 (2.7 – 8.3)	0.1 (0.001 – 0.3)
irregular border	0.84 (0.74 – 0.93)	0.31	80 (24/30) (61 – 92)	87 (78/90) (78 – 93)	6.0 (3.1 – 14.8)	0.2 (0.1 – 0.5)
cystic endometrium	0.84 (0.75 – 0.93)	0.27	77 (23/30) (58 – 90)	86 (77/90) (77 – 92)	5.3 (2.7 – 12.5)	0.3 (0.1 – 0.5)
hyperechogenic endometrium	0.81 (0.71 – 0.91)	0.31	70 (21/30) (51 – 85)	87 (78/90) (78 – 93)	5.3 (2.6 – 13.1)	0.3 (0.1 – 0.6)
Endometrial thickness and						
irregular branching[2]	0.86 (0.77 – 0.95)	0.29	77 (23/30) (58 – 90)	88 (79/90) (79 – 94)	6.3 (3.0 – 15.9)	0.3 (0.1 – 0.5)
areas of densely packed vessels[3]	0.85 (0.76 – 0.94)	0.28	83 (25/30) (65 – 94)	82 (74/90) (73 – 89)	4.7 (2.6 – 9.7)	0.2 (0.03 – 0.4)
color ´splashes´	0.83 (0.74 – 0.92)	0.33	77 (23/30) (58 – 90)	82 (74/90) (73 – 89)	4.3 (2.3 – 9.1)	0.3 (0.1 – 0.5)
many vessels	0.82 (0.72 – 0.91)	0.24	80 (24/30) (61 – 92)	77 (69/90) (67 – 85)	3.4 (2.0 – 6.5)	0.3 (0.1 – 0.5)
pattern A**	0.81 (0.72 – 0.91)	0.26	77 (23/30) (58 – 90)	77 (69/90) (67 – 85)	3.3 (1.9 – 6.2)	0.3 (0.1 – 0.6)
Endometrial thickness and heterogenous echogenicity and						
areas of densely packed vessels[4]	0.92 (0.86 – 0.98)	0.28	83 (25/30) (65 – 94)	89 (80/90) (81 – 95)	7.5 (3.7 – 20.1)	0.2 (0.03 – 0.4)
irregular branching	0.92 (0.85 – 0.98)	0.15	87 (26/30) (69 – 96)	83 (75/90) (74 – 90)	5.2 (2.9 – 11.0)	0.2 (0.02 – 0.4)

Continued

Table S2 continued

Model	ROC area Estimate (95% CI)	Optimal probability cutoff	Sensitivity, %, (n) (95% CI)	Specificity %, (n) (95% CI)	LR+ (95% CI)	LR- (95% CI)
color 'splashes'	0.91 (0.85 – 0.97)	0.31	80 (24/30) (61 – 92)	90 (81/90) (82 – 95)	8.0 (3.8 – 23.3)	0.2 (0.1 – 0.4)
Endometrial thickness and irregular border and						
areas of densely packed vessels	0.86 (0.77 – 0.95)	0.44	73 (22/30) (54 – 88)	92 (83/90) (85 – 97)	9.4 (3.9 – 33.1)	0.3 (0.1 – 0.5)
color 'splashes'	0.85 (0.76 – 0.94)	0.52	67 (20/30) (47 – 83)	96 (86/90) (89 – 99)	15.0 (4.8 – 107.6)	0.3 (0.1 – 0.6)
Heterogenous echogenicity and						
areas of densely packed vessels[5]	0.89 (0.81 – 0.97)	0.41	73 (22/30) (54 – 88)	92 (83/90) (85 – 97)	9.4 (3.9 – 33.1)	0.3 (0.1 – 0.5)
color 'splashes'	0.87 (0.78 – 0.96)	0.42	73 (22/30) (54 – 88)	92 (83/90) (85 – 97)	9.4 (3.9 – 33.1)	0.3 (0.1 – 0.5)
irregular branching	0.87 (0.78 – 0.96)	0.23	80 (24/30) (61 – 92)	88 (79/90) (79 – 94)	6.5 (3.3 – 16.7)	0.2 (0.1 – 0.5)
branching	0.86 (0.78 – 0.95)	0.32	73 (22/30) (54 – 88)	92 (83/90) (85 – 97)	9.4 (3.9 – 33.1)	0.3 (0.1 – 0.5)
pattern A**	0.86 (0.77 – 0.95)	0.34	73 (22/30) (54 – 88)	92 (83/90) (85 – 97)	9.4 (3.9 – 33.1)	0.3 (0.1 – 0.5)
Irregular endometrial-myometrial border and						
areas of densely packed vessels	0.80 (0.69 – 0.90)	0.56	57 (17/30) (37 – 75)	97 (87/90) (91 – 99)	17.0 (4.4 – 209.2)	0.4 (0.2 – 0.7)
color 'splashes'	0.78 (0.67 – 0.89)	0.55	50 (15/30) (31 – 69)	98 (88/90) (92 – 100)	22.5 (4.6 – 868.4)	0.5 (0.3 – 0.7)

Table S2 continued

Model	ROC area Estimate (95% CI)	Optimal probability	Sensitivity, %, (n) (95% CI)	Specificity %, (n) (95% CI)	LR+ (95% CI)	LR- (95% CI)
branching	0.77 (0.66 – 0.87)	0.40	63 (19/30) (44 – 80)	88 (79/90) (79 – 94)	5.2 (2.4 – 13.9)	0.4 (0.2 – 0.6)
many vessels	0.75 (0.64 – 0.86)	0.39	67 (20/30) (47 – 83)	84 (76/90) (75 – 91)	4.3 (2.1 – 10.0)	0.4 (0.2 – 0.7)
pattern A**	0.75 (0.63 – 0.86)	0.40	67 (20/30) (47 – 83)	87 (78/90) (78 – 93)	5.0 (2.4 – 12.7)	0.4 (0.2 – 0.6)

LR, likelihood ratio; ROC, receiver operating characteristic; CI, confidence interval; *Values above the probability cut off indicate malignancy; **Vascular pattern A according to Alcazar et al[4]

Probability of malignancy is calculated as $[e^Z/1+e^Z]$ where e = 2.718 (mathematical constant and base value of natural logarithms) and z is calculated for each logistic regression model

[1] z = -3.988 + (0.125 x endometrial thickness in mm) + (3.012 x echogenicity coded 0 if homogeneous and 1 if heterogeneous)
[2] z = -4.151 + (0.156 x endometrial thickness in mm) + (2.886 x branching coded 0 if regular and 1 if irregular)
[3] z = -4.411 + (0.167 x endometrial thickness in mm) + (2.151 x areas of densely packed vessels coded 0 if absent and 1 if present)
[4] z = -4.462 + (0.113 x endometrial thickness in mm) + (2.832 x echogenicity coded 0 if homogeneous and 1 if heterogeneous) + (1.907 x areas of densely packed vessels coded 0 absent and 1 if present)
[5] z = -3.037 + (3.243 x echogenicity coded 0 if homogeneous and 1 if heterogeneous) + (2.050 x areas of densely packed vessels coded 0 if absent and 1 if present)

II

Prediction of endometrial malignancy in women with postmenopausal bleeding and sonographic endometrial thickness ≥4.5 mm

Gina Opolskiene, MD, Povilas Sladkevicius, MD, PhD, Lil Valentin, MD, PhD

Department of Obstetrics and Gynecology, Skåne University Hospital Malmö, Lund University, Malmö, Sweden

Short title: endometrial cancer

Key words: Endometrial neoplasms; postmenopause; ultrasonography; Doppler, color

Correspondence: Gina Opolskiene, Department of Obstetrics and Gynecology, Skåne University Hospital Malmö, S-20502 Malmö, Sweden

Telephone: +46 40 332094 Fax: +46 40 962600

Email: **Gina.Opolskiene@med.lu.se**

Abstract

Objectives To build mathematical models to evaluate the individual risk of endometrial malignancy in women with postmenopausal bleeding and sonographic endometrial thickness ≥4.5 mm using clinical data, sonographic endometrial thickness and power Doppler ultrasound findings.

Methods Of 729 consecutive patients with postmenopausal bleeding, 261 with sonographic endometrial thickness ≥4.5 mm and no fluid in the uterine cavity were included. They underwent transvaginal two-dimensional gray scale and power Doppler ultrasound examination of the endometrium. The ultrasound image showing the most vascularized section through the endometrium as assessed by power Doppler was frozen, the endometrium was outlined, and the percentage vascularized area (vascularity index) was calculated using computer software. The ultrasound examiner also estimated the color content of the endometrial scan on a visual analogue scale (VAS) graded from 0 to 100 (VAS score). A structured history was taken to collect clinical information. Multivariate logistic regression was used to create mathematical models to predict endometrial malignancy.

Results There were 63 (24%) malignant and 198 benign endometria. Women with malignant endometrium were older (median 74 years vs. 65; P = 0.0005) and fewer used hormonal replacement therapy (HRT) and warfarin. Women with malignant endometrium had thicker endometrium (median 20.8 mm vs. 10.2; P = 0.0005) and higher values for vascularity index and VAS score. When using only clinical data to build a model to estimate the risk of endometrial malignancy, a model including the variables age, use of warfarin and use of HRT had the largest area (0.74, 95% CI 0.67 – 0.81) under the receiver operating characteristic curve (AUC). A model including age, use of warfarin and endometrial thickness had AUC (0.82, 95% CI 0.76 – 0.87), and one including age, use of HRT, endometrial thickness and vascularity index had AUC (0.91, 95% CI 0.87 – 0.95). Using a risk cut-off of 11%, the model including age, use of HRT, endometrial thickness and vascularity index had sensitivity 90%, specificity 71%, positive likelihood ratio 3.14 and negative likelihood ratio 0.13.

Conclusions The diagnostic performance of models predicting endometrial cancer increases substantially when sonographic endometrial thickness and power Doppler information are added to clinical variables. The models are likely to be clinically useful but need to be prospectively validated.

Introduction

The measurement of sonographic endometrial thickness makes it possible to separate women with postmenopausal bleeding into a low risk group and a high risk group: if the sonographic endometrial thickness is less than 5 mm the risk of endometrial malignancy is low, if it is 5 mm or more, the risk of endometrial malignancy is high [1]. Women with postmenopausal bleeding and thick endometrium must undergo endometrial sampling. Endometrial sampling using a simple endometrial outpatient sampling device is a simple procedure. However, many postmenopausal women have a stenotic cervical canal making outpatient endometrial sampling impossible [2,3]. Moreover, endometrial samples obtained by an outpatient endometrial sampling device often contain insufficient material, and endometrial cancers may be missed [4]. If outpatient sampling fails, the woman needs to undergo dilatation and curettage (D&C) or hysteroscopy in anaesthesia, or possibly in regional analgesia. However, a substantial proportion of postmenopausal women are at high operative risk making D&C or hysteroscopy in anesthesia inadvisable unless absolutely necessary. It would be useful to have access to a method to estimate the individual risk of malignancy in women with postmenopausal bleeding and thick endometrium, because this would make it possible to individualize and optimize management. Clinical factors that increase the risk of endometrial malignancy are age, overweight, diabetes, hypertension, and nulliparity [5]. Other factors that affect the risk are endometrial thickness and vascularity as determined with ultrasound [6,7], use of hormone replacement therapy [1,7] or anticoagulants [5].

The purpose of this study was to build mathematical models to calculate the risk of endometrial malignancy in women with postmenopausal bleeding and endometrial thickness ≥4.5 mm (rounded up to 5 mm) using clinical data, sonographic endometrial thickness and power Doppler ultrasound findings.

Methods

Since 2002, women consulting for postmenopausal bleeding in our hospital are managed in a Postmenopausal Bleeding Clinic run by two gynecologists specialized in gynecological ultrasound (PS and LV). A woman is considered to be postmenopausal if she reports absence of menstruation for at least 1 year after the age of 40 years provided that the amenorrhea is not explained by pregnancy, medication or disease. Postmenopausal bleeding is defined as any vaginal bleeding in a postmenopausal woman not on hormone replacement therapy, or unscheduled vaginal bleeding in a postmenopausal woman on hormonal replacement therapy. Women managed in our Postmenopausal Bleeding Clinic undergo a transvaginal ultrasound examination with measurement of endometrial thickness. The ultrasound examinations are carried out when the women present without any specific timing with regard to the "cycle" in women on sequential hormone replacement therapy. If the endometrial thickness is $\leq$4.4 mm, the woman is dismissed but asked to come back if she bleeds again. If the endometrium measures $\geq$4.5 mm (rounded up to 5 mm), saline contrast sonohysterography (SCHS), i.e. instillation of saline into the uterine cavity during scanning, is performed as described earlier [8]. If there are no focal lesions in the uterine cavity at SCSH, an endometrial sample using Endorette® is taken, or D&C is suggested to the referring physician. The Endorette (Medscand AB, Malmö, Sweden) is a sterile curette with a polyethylene piston which slides within a straight but flexible polypropylene sheath with four lateral holes near its tip. Its length is 285 mm and its outer diameter is 2.6 mm. If there are focal lesions at SCSH, or if SCSH fails, hysteroscopic resection of focal lesions/endometrium is recommended. In this work (as well as in all our previous work), a focal lesion is defined as any protrusion into the endometrial cavity above the surrounding endometrium.

Women with endometrium $\geq$4.5 mm and no fluid in the uterine cavity at the transvaginal scan were eligible for inclusion in our study. Exclusion criteria were absence of postprocessed ultrasound images for computer analysis or absence of a reliable histological diagnosis.

Women who were eligible and accepted participation were examined as described below. A structured history was taken following a standardized research protocol regarding age at menopause, parity, hormone replacement therapy, weight, height, hypertension, diabetes, and current use of anticoagulants (warfarin, clopidogrel bisulfate, ticlopidine, low dose acetyl salicylic acid). The question about anticoagulants was included in our research protocol in 2007. For patients examined before 2007 the information on use of anticoagulants at the time of the ultrasound examination was collected retrospectively from patient records.

Transvaginal sonography was carried out by one of two examiners (LV or PS) using a Sequoia 512 ultrasound system (Siemens Medical Solutions Inc., Ultrasound Division, Mountain View, CA) equipped with a 5 – 8 MHz transvaginal transducer. All women were examined in the lithotomy position with an empty bladder. First, conventional gray scale ultrasound examination of the uterus was performed. The image was magnified to contain only the uterine corpus, and power Doppler ultrasound examination of the endometrium was carried out using predetermined, standardized settings (frequency 6 MHz; power Doppler gain 50; dynamic range 10 dB; edge 1; persistence 2; color map 1; gate 2; filter 3). To detect the most vascularized section through the endometrium, the corpus uteri was scanned with power Doppler ultrasound in the sagittal plane from one side to the other. The image of the most vascularized section through the endometrium, as estimated subjectively, was frozen, and the endometrium was outlined with callipers using the trace function of the ultrasound system. Then the image was postprocessed, i.e. the gray-scale echoes were removed, so that the colored power Doppler pixels were shown on a black background (Figure 1) as previously described [7]. While scanning the endometrium in the power Doppler mode, the ultrasound examiner also estimated the color content of the endometrial scan on a visual analogue scale (VAS) graded from 0 to 100 (VAS score).

The postprocessed images were stored electronically and transferred to an offline computer where they were transformed into Tag Image File (TIF) format. They were analyzed by a single examiner (GO) using dedicated software to quantify the color content of the endometrial scan. The software was developed using MATLAB 6.0TM software (TheMathWorks Inc., Natick, MA, USA) [9]. When analysing the images, the outline of the endometrium was drawn again following the previously drawn line in the postprocessed image. The software then calculated the area of the endometrium, the area of color pixels and the ratio between the two areas expressed as a percentage (vascularity index - VI).

Clinical and ultrasound information were compared with the final histological diagnosis of the endometrium, the final histological diagnosis being obtained by D&C or hysteroscopic resection (the indications for D&C and hysteroscopy have been described above) or by hysterectomy. Staging of malignant tumors was done by the attending physician in accordance with the classification system recommended by the International Federation of Gynecology and Obstetrics [10], but lymphadenectomy was performed only in high risk cases.

To determine intra-observer repeatability and reliability [11] of calculations of vascularity index we used the images of 20 patients. The 20 patients were selected from our statistical data sheet so as to include 10 consecutive patients with malignant endometrium and 10 consecutive

patients with benign endometrium. The 20 images were analyzed twice with regard to vascularity index by one examiner (GO) 6 months apart.

To determine intra-observer repeatability and reliability and inter-observer reproducibility and reliability [11] of the VAS score we used the images of 30 patients. These patients were selected from our statistical data sheet so as to include 10 consecutive patients with malignant endometrium (the same as those included for determination of intra-observer repeatability and reliability of the vascularity index) and 20 consecutive patients with benign endometrium (the same 10 patients as those included for determination of intra-observer repeatability and reliability of the vascularity index plus the subsequent consecutive 10 patients with benign endometrium). The 30 images were analyzed with regard to the VAS score by two independent observers blinded to each others results (LV and PS). Each of the two examiners analyzed each image twice, at least 2 weeks apart. When determining the inter-observer reproducibility and reliability we used the results of the first analysis of each observer.

Statistical calculations were undertaken using the Statistical Package for the Social Sciences (SPSS Inc., Chicago, Illinois, USA, version 16.0). The statistical significance of a possible relationship of endometrial malignancy with clinical variables (age, body mass index, diabetes, hypertension, nulliparity, use of hormone replacement therapy, anticoagulants, or levothyroxine) and ultrasound variables (sonographic endometrial thickness, vascularity index, VAS score) was determined using univariate logistic regression with the likelihood ratio test. Two-tailed P-values <0.05 were considered statistically significant. Multivariate logistic regression was used to build models to predict malignancy. To avoid overfitting a maximum of six predicting variables were allowed in a model, the likelihood ratio test yielding a P <0.05 being the criterion for including a variable in a model. Receiver operating characteristic (ROC) curves were drawn for single predicting clinical and ultrasound variables and for the multivariate logistic regression models. The area under the ROC curve and the 95% confidence interval (CI) of this area were calculated. If the lower limit of the CI for the area under the ROC curve was >0.5, the variable/model was considered to have discriminatory potential. The ROC curves were also used to determine the mathematically best cut-off value to predict malignancy for continuous single variables as well as for logistic regression models, the mathematically best cut-off value being defined as that corresponding to the point on the ROC curve situated farthest from the reference line. The sensitivity, specificity, positive and negative likelihood ratios (LR) of the mathematically best cut-off value and their 95% confidence intervals were then calculated. We also calculated the specificity at 90% sensitivity

for each model. We defined the best diagnostic test as the one with the largest area under the ROC curve and the highest specificity at 90% sensitivity.

Intra-observer repeatability and inter-observer reproducibility was estimated as the difference between two measurement results. The differences between the measured values were plotted against the mean of the two measurements (Bland-Altman plots) to assess the relationship between the differences and the magnitude of the measurements [12]. Systematic bias between two analyses was estimated by calculating the 95% CI of the mean difference (mean difference +/- 2 SE, standard error of the mean). If zero lay within this interval no bias was assumed to exist between the two measurements. Intra-observer repeatability was expressed as the repeatability coefficient, the absolute difference between two measurements on a subject being expected to differ by no more than the repeatability coefficient on 95% of occasions [11]. Inter-observer reproducibility was expressed as mean difference and limits of agreement [11]. Intra- and inter-observer reliability was determined by calculating the intra-class correlation coefficient (ICC) using analysis of variance (two way random model - absolute agreement; this allows generalization of the results to a population of observers) [11]. The ICC indicates the proportion of the total variance in measurement results that can be explained by differences between the individuals examined. The more variable the population investigated, the greater the ICC and the less variable the population, the smaller the ICC [11].

The Ethics Committee of Lund University approved the study protocol. Informed consent was obtained from all participants, after the nature of the procedures had been fully explained.

Results

Between November 2002 and June 2009, 729 women with postmenopausal bleeding were examined in our Postmenopausal Bleeding Clinic. For women who were examined more than once because of recurrent bleeding (n = 40), the examination where the endometrium measured ≥4.5 mm was included. If it measured ≥4.5 mm on more than one occasion (n = 18), the examination with the most complete information was used in our statistical analysis.

Of the 729 patients with postmenopausal bleeding, 371 were eligible for inclusion in our study. None declined to participate but 110 had to be excluded (Figure 2). Thus, 261 patients were used in our statistical analysis. Demographic characteristics of the women included and excluded are shown in Table 1. The included and excluded patients had similar clinical background characteristics, but the women excluded had thinner endometrium (median 8 mm versus 12 mm) and fewer had endometrial malignancy (6% vs. 24 %).

Of the 261 women included, 198 had benign and 63 had malignant endometrium. There were eight cases of rare type of endometrial cancer (Table 1). Women with malignant endometrium were older (74 years vs. 65; P = 0.0005) than women with benign endometrium, fewer used hormonal replacement therapy (14% vs. 35%; P = 0.001) and fewer used warfarin (2% vs. 12%; P = 0.005). Women with endometrial malignancy had thicker endometrium than those with benign endometrium (median 20.8 mm vs 10.2; P = 0.0005) and higher percentage vascularized area of the endometrium as assessed both objectively and subjectively (median vascularity index 43.2% vs. 13.6%, P = 0.0005; median VAS score 65 vs. 13; P = 0.0005) (Table 2). According to the area under the ROC curve, age, the use of hormonal replacement therapy, sonographic endometrial thickness, vascularity index and VAS score had potential to discriminate between benign and malignant endometrium.

Results of univariate and multivariate logistic regression (odds ratios) are shown in Table 3, and the diagnostic performance of single variables and the logistic regression models created is shown in Table 4. No more than four variables entered any logistic regression model. Models including only clinical variables had poor diagnostic performance, the best of these models having an AUC of 0.74. Adding endometrial thickness to clinical variables substantially improved the diagnostic performance, the best model including clinical variables and endometrial thickness having an AUC of 0.82. Adding both endometrial thickness and Doppler results to the clinical variables improved model performance even further, the best model including clinical variables, endometrial thickness and Doppler results having an AUC of 0.91. ROC curves are shown in Figure 3. The model with the largest area under the ROC

curve (0.91) and the highest specificity at 90% sensitivity included the variables endometrial thickness, vascularity index, age, and use of hormone replacement therapy. The probability of malignancy for this model is calculated as $[e^z/(1+e^z)]$ where e = 2.718 (base value of natural logarithms) and z is calculated as follows: z = -9.356 + (0.084 x endometrial thickness in mm) + (0.088 x vascularity index in percent) + (0.069 x age in years) + (-1.333 x use of HRT; coded 1 if used and 0 if not used). Using a risk cut-off of 0.11 (i.e., risk 11%) to indicate malignancy, this model had sensitivity 90.5%, specificity 71%, LR+ 3.14 and LR- 0.13.

The results of both univariate and multivariate analyses were almost identical when we performed these analyses including also the 50 patients with unreliable histological diagnosis (unreliable histological diagnosis being defined as insufficient endometrial material, diagnosis made on the basis of an Endorette® sample only, final diagnosis made on the basis of a D&C sample despite there being focal lesions in the uterine cavity, bicornuate uterus where it was not clear from which of the two uterine cavities the specimens had been taken, see Figure 2).

Results with regard to intra-observer repeatability and reliability for analysis of vascularity index are shown in Table 5. There was no systematic difference between the first and second analysis by the same observer, the differences did not change with the magnitude of the measurement results (Bland Altman plots), the difference between two measurements was small and reliability was excellent (ICC 1.00).

Results with regard to intra-observer repeatability and reliability and inter-observer reproducibility and reliability for evaluation of the VAS score are shown in Table 6. The differences between two measurements did not change with the magnitude of the measurement results (Bland Altman plots). One observer systematically obtained higher values at the second analysis, while the other observer systematically obtained lower values at the second analysis. Intra- and inter-observer reliability was good (all ICC values >0.95).

Discussion

The results of our study show that the diagnostic performance of logistic regression models constructed to predict endometrial malignancy in women with postmenopausal bleeding and sonographic endometrial thickness ≥4.5 mm increases substantially when adding sonographic endometrial thickness and power Doppler ultrasound information to clinical variables. The best multivariate logistic regression model for calculating the individual risk of endometrial malignancy included the variables age, use of hormone replacement therapy, endometrial thickness and vascularity index. This model had an AUC of 0.91 and a specificity of 71% at 90% sensitivity corresponding to LR+ 3.14 and LR- 0.13, i.e. this model was fairly good at excluding malignancy.

It is a strength of our study that we included only women with a reliable histological diagnosis. We excluded all cases where there was insufficient endometrial material for histological diagnosis, all cases where the diagnosis was made on the basis of an Endorette® sample only, and all cases where the final diagnosis was made on the basis of a D&C sample despite there being focal lesions in the uterine cavity. Most other studies where logistic regression models were created to calculate the risk of endometrial malignancy included also women where the diagnosis was made on the basis of an endometrial sample taken by an outpatient endometrial sampling device [13,14], and one study included even women where no endometrial sample was taken [5]. We prefer having a study population that is well defined with regard to histopathological diagnosis to having a larger but less well defined study population. However, we obtained almost identical results when we performed our analyses without making the exclusions described above. It is also a strength of our study that we determined the reproducibility of the VAS score and the vascularity index.

The limitations of our study are that information on use of anticoagulants was collected retrospectively from patient records for some of the women and that we did not include acquisition of color Doppler images in our analysis of the reproducibility/repeatability and reliability of the vascularity index and VAS score. It is also a limitation that our results are only applicable to ultrasound examinations carried out using the same ultrasound system and transducer that we used in the current study. This, unfortunately, is a limitation that applies to all studies where color and power Doppler variables are involved, since the color content of a color or power Doppler scan depends heavily on the Doppler sensitivity of the ultrasound system used.

Others, too, have tried to create multivariate logistic regression models to calculate the

individual risk of malignancy in women with postmenopausal bleeding [5-7,13,14]. Some models were built exclusively for women not on hormone replacement therapy and included also women with sonographic endometrial thickness ≤4.4 mm [5,13,14]. Predicting variables in these models were endometrial thickness [5,13,14], age [5,13], body mass index [5,13], nulliparity [5], diabetes [5] (all increasing the risk of malignancy), and use of anticoagulants [5] (decreasing the risk of malignancy). In agreement with our results, Opmeer et al found that adding information on endometrial thickness to clinical information substantially increased the diagnostic performance of risk calculation models. The AUC of their model increased from 0.76 to 0.90 when they added endometrial thickness to clinical information [5]. In the study by Randelzhofer et al, endometrial thickness, irregular endometrial echogenicity and irregular endometrial-myometrial border increased the risk of malignancy [14].

We decided to build models only for women with endometrial thickness ≥4.5 mm for two reasons. First, women with postmenopausal bleeding and endometrial thickness <4.4 mm are at low risk of endometrial cancer. Even though some have suggested that a cut-off level of 3 mm should be used to exclude endometrial carcinoma in women with postmenopausal bleeding [15] and have emphasized that women with postmenopausal bleeding and thin endometrium do have 1% risk of endometrial cancer [16], there is general agreement that it is safe to refrain from endometrial sampling if endometrial thickness is <5 mm [1,17-19]. Second, because we wanted to determine the value of adding information on Doppler ultrasound results to clinical information and gray scale ultrasound results, we needed to exclude women with thin endometrium, because it is almost impossible to evaluate Doppler ultrasound findings in endometria no thicker than 4.4 mm. We found it logical to build models including also women on hormone replacement therapy. Indeed, use of hormone replacement therapy proved to be an important variable in our models: it substantially decreased the odds of malignancy. This is in agreement with the results of another study where models were built to estimate the risk of endometrial malignancy in women with postmenopausal bleeding and endometrial thickness ≥4.5 mm [7]. On the other hand, in another study on women with postmenopausal bleeding and endometrial thickness ≥4.5 mm, hormone replacement therapy did not enter a logistic regression model to predict malignancy [6]. In both studies cited [6,7] endometrial thickness and heterogeneous internal echogenicity of the endometrium increased the risk of malignancy, and in both studies adding information on the results of Doppler examination of the endometrium improved the performance of the logistic regression models. The performance of the models created in the two studies cited [6,7] was similar to that of the best model in our study (AUC 0.88 [7] and 0.92 [6] versus 0.91 in our current study). Some might want to argue that it was unnecessary to build

yet another logistic regression model to estimate the risk of endometrial malignancy in women with postmenopausal bleeding and endometrial thickness ≥4.5 mm, because several models were already available for this purpose [6,7]. However, some of the models published would require highly experienced ultrasound examiners to evaluate the internal echogenicity of the endometrium [6,7] and the morphology of endometrial vessels [6]. Moreover, the simplest previously published model was built on a rather small number of cases and included only three variables (endometrial thickness, the color content of the endometrial scan at color Doppler examination determined using computer software, and use of hormone replacement therapy) [7]. We wanted to collect a larger number of cases to have the possibility to include more variables in our model, and we wanted to create a model that would only include ultrasound variables on which one can obtain information even if one is only a moderately experienced ultrasound examiner.

We believe that our models including clinical, gray scale and Doppler ultrasound findings are potentially clinically useful. The estimated risk calculated by the models is meant to help clinicians optimize the management of their patients. For example, it is clinically important to know when sampling the endometrium is unnecessary, if invasive procedures involving anaesthesia are necessary to get a representative sample and the patient is a great operative risk. It is also clinically important to know which patients should be given priority to invasive procedures involving anaesthesia if there are long waiting lists for such procedures and such procedures are necessary to obtain a representative sample.

Unfortunately, the method of objectively calculating the color content of the endometrial scan offline is too cumbersome for clinical use. On the other hand, the ultrasound system that we used allows this to be done during scanning using the software of the ultrasound system. The software is simple and any ultrasound company could install such software in their ultrasound systems. To estimate the color content of the endometrial scan subjectively is quick and easy, and our results suggest that the results of subjective evaluation of the color content of the endometrial scan are reproducible and reliable. However, our model including the VAS score did not perform as well as our model including the objectively calculated vascularity index. Prospective validation will show which of our two models including Doppler information will work best in clinical practice.

Acknowledgements

References

1. Smith-Bindman R, Kerlikowske K, Feldstein VA, Subak L, Scheidler J, Segal M, Brand R,Grady D. Endovaginal ultrasound to exclude endometrial cancer and other endometrial abnormalities. *JAMA* 1998; **280**: 1510-1517.

2. De Silva BY, Stewart K, Steven JD, Sathanandan M. Transvaginal ultrasound measurement of endometrial thickness and endometrial pipelle sampling as an alternative diagnostic procedure to hysteroscopy and dilatation and curettage in the management of post-menopausal bleeding. *J Obstet Gynaecol* 1997; **17**: 399-402.

3. Epstein E, Skoog L, Valentin L. Comparison of Endorette and dilatation and curettage for sampling of the endometrium in women with postmenopausal bleeding. *Acta Obstet Gynecol Scand* 2001; **80**: 959-964.

4. Goldstein SR. The role of transvaginal ultrasound or endometrial biopsy in the evaluation of the menopausal endometrium. *Am J Obstet Gynecol* 2009; **201**: 5-11.

5. Opmeer BC, van Doorn HC, Heintz AP, Burger CW, Bossuyt PM, Mol BW. Improving the existing diagnostic strategy by accounting for characteristics of the women in the diagnostic work up for postmenopausal bleeding. *BJOG* 2007; **114**: 51-58.

6. Opolskiene G, Sladkevicius P, Valentin L. Ultrasound assessment of endometrial morphology and vascularity to predict endometrial malignancy in women with postmenopausal bleeding and sonographic endometrial thickness >or= 4.5 mm. *Ultrasound Obstet Gynecol* 2007; **30**: 332-340.

7. Epstein E, Skoog L, Isberg PE, De Smet F, De Moor B, Olofsson PA, Gudmundsson S, Valentin L. An algorithm including results of gray-scale and power Doppler ultrasound examination to predict endometrial malignancy in women with postmenopausal bleeding. *Ultrasound Obstet Gynecol* 2002; **20**: 370-376.

8. Epstein E, Valentin L. Managing women with post-menopausal bleeding. *Best Pract Res Clin Obstet Gynaecol* 2004; **18**: 125-143.

9. Marret H, Sauget S, Giraudeau B, Body G, Tranquart F. Power Doppler vascularity index for predicting malignancy of adnexal masses. *Ultrasound Obstet Gynecol* 2005; **25**: 508-513.

10. Shepherd JH. Revised FIGO staging for gynaecological cancer. *Br J Obstet Gynaecol* 1989; **96**: 889-892.

11. Bartlett JW, Frost C. Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. *Ultrasound Obstet Gynecol* 2008; **31**: 466-475.

12. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; **1**: 307-310.

13. Ferrazzi E, Torri V, Trio D, Zannoni E, Filiberto S, Dordoni D. Sonographic endometrial thickness: a useful test to predict atrophy in patients with postmenopausal bleeding. An Italian multicenter study. *Ultrasound Obstet Gynecol* 1996; **7**: 315-321.

14. Randelzhofer B, Prompeler HJ, Sauerbrei W, Madjar H, Emons G. Value of sonomorphological criteria of the endometrium in women with postmenopausal bleeding: a multivariate analysis. *Ultrasound Obstet Gynecol* 2002; **19**: 62-68.

15. Timmermans A, Opmeer BC, Khan KS, Bachmann LM, Epstein E, Clark TJ, Gupta JK, Bakour SH, van den Bosch T, van den Doorn HC, Cameron ST, Giusa MG, Dessole S, Dijkhuizen FP, Ter Riet G, Mol BW. Endometrial thickness measurement for detecting endometrial cancer in women with postmenopausal bleeding: a systematic review and meta-analysis. *Obstet Gynecol* 2010; **116**: 160-167.

16. Timmermans A, Opmeer BC, Veersema S, Mol BW. Patients' preferences in the evaluation of postmenopausal bleeding. *BJOG* 2007; **114**: 1146-1149.

17. Gull B, Carlsson S, Karlsson B, Ylostalo P, Milsom I, Granberg S. Transvaginal ultrasonography of the endometrium in women with postmenopausal bleeding: is it always necessary to perform an endometrial biopsy? *Am J Obstet Gynecol* 2000; **182**: 509-515.

18. Goldstein RB, Bree RL, Benson CB, Benacerraf BR, Bloss JD, Carlos R, Fleischer AC, Goldstein SR, Hunt RB, Kurman RJ, Kurtz AB, Laing FC, Parsons AK, Smith-Bindman R, Walker J. Evaluation of the woman with postmenopausal bleeding: Society of Radiologists in Ultrasound-Sponsored Consensus Conference statement. *J Ultrasound Med* 2001; **20**: 1025-1036.

19. Gupta JK, Chien PF, Voit D, Clark TJ, Khan KS. Ultrasonographic endometrial thickness for diagnosing endometrial pathology in women with postmenopausal bleeding: a meta-analysis. *Acta Obstet Gynecol* Scand 2002; **81**: 799-816.

20. Epstein E, Skoog L, Valentin L. Comparison of Endorette and dilatation and curettage for sampling of the endometrium in women with postmenopausal bleeding. *Acta Obstet Gynecol Scand* 2001; **80**: 959-964.

a

b

c
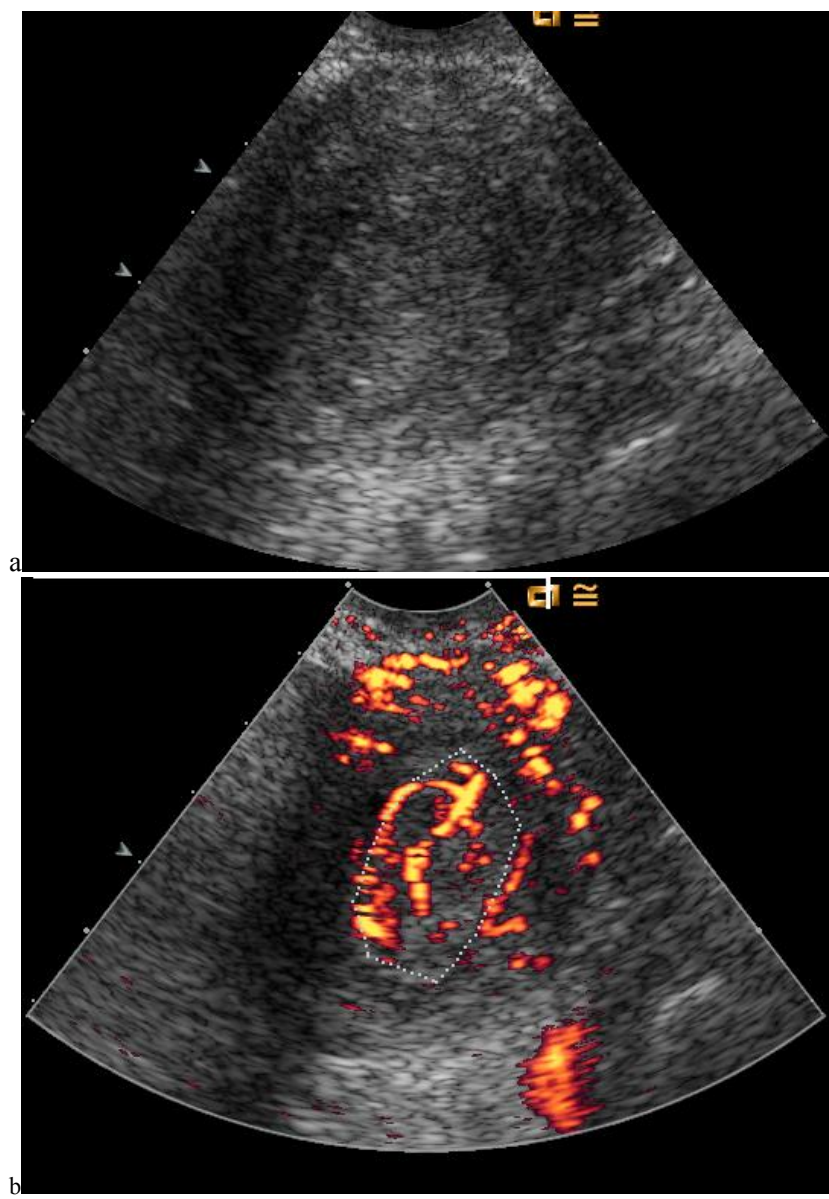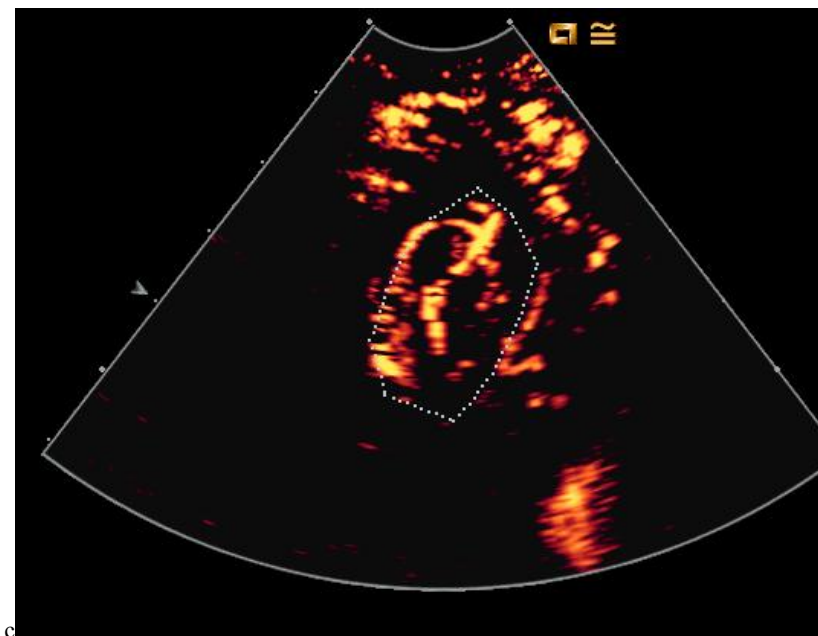
Figure 1. Ultrasound images of the endometrium illustrating how the ultrasound images were processed for computer analysis of the color content of the endometrial scan. Gray scale image of the endometrium (a), power Doppler image of the same endometrium where the outline of the endometrium has been traced using the software of the ultrasound system (b), gray scale echoes have been removed using the software of the ultrasound system so that the power Doppler pixels are shown against a black background (c). Image (c) was used to quantify the color content of the endometrial scan using dedicated software. The histopathological diagnosis of this case was carcinosarcoma.
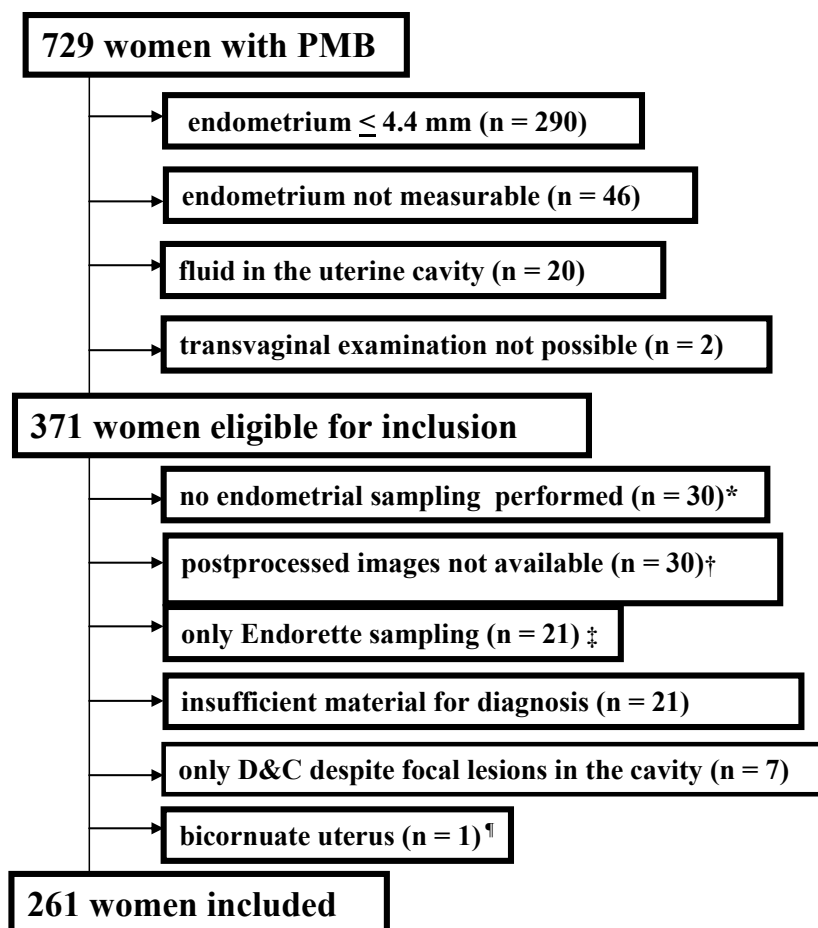
Figure 2. Flow chart describing patient selection. Of 729 women with postmenopausal bleeding (PMB), 371 were eligible for inclusion, but 110 needed to be excluded

* some women refused any kind of endometrial sampling, some moved to another country or city and information on endometrial histology is unavailable, some were too ill to undergo sampling under anaesthesia, and some died before an endometrial specimen was obtained

† the Sequoia ultrasound system was not available, or the ultrasound examiner forgot to save postprocessed ultrasound images

‡ the reason for excluding samples obtained only by Endorette® was that we wanted to be sure that only reliable samples were used to establish the final diagnosis [3,19]

¶It was not clear from which of the two uterine cavities the specimens had been taken.
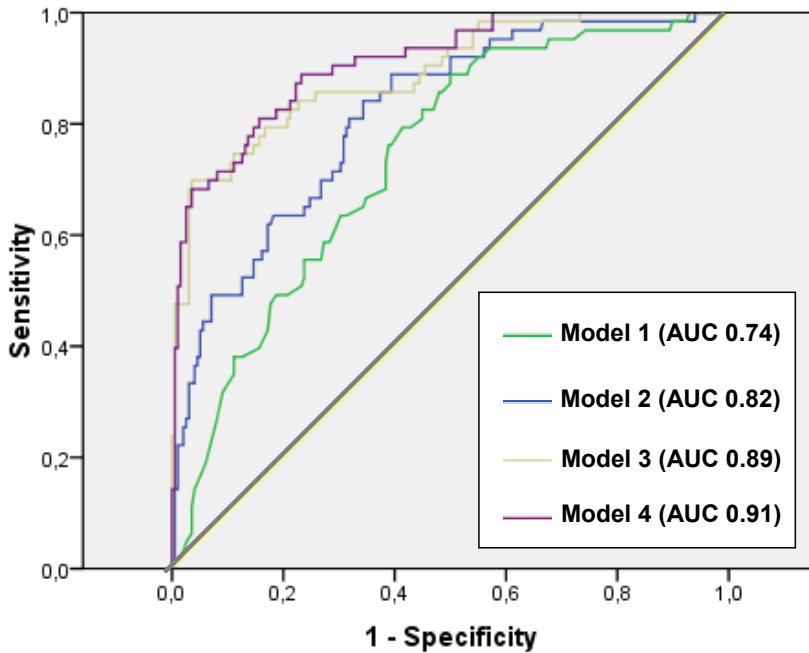
Figure 3. Receiver operating characteristic (ROC) curves of logistic regression models to calculate the risk of endometrial malignancy in women with postmenopausal bleeding and sonographic endometrial thickness ≥4.5 mm. AUC, area under the ROC curve. Model 1 includes the variables age, use of warfarin and use of hormone replacement therapy. Model 2 includes the variables age, use of warfarin and sonographic endometrial thickness. Model 3 includes the variables age, hormone replacement therapy, sonographic endometrial thickness and VAS score. Model 4 includes the variables age, hormone replacement therapy, sonographic endometrial thickness and vascularity index.

Table 1 Characteristics of included and excluded patients

	Included (n = 261)	Excluded (n = 110)
Age, years; median, range	67 (47 – 91)	69 (42 – 93)
BMI, (kg/m^2); median, range	28 (17 – 50)*	25 (17 – 55)†
Nulliparous, n (%)	31 (12)‡	15 (14)§
Hypertension, n (%)	111 (43)	41 (40)¶
Diabetes, n (%)	35 (13)	9 (9)¶
Hormone replacement therapy, n (%)	79 (30)	39 (37)§
Use of anticoagulants, n (%)	63 (24)	31 (28)
Use of levothyroxine, n (%)	32 (12)	16 (15)
Endometrial thickness, mm; median, range	12.0 (4.5 – 75.6)	8.3 (4.5 – 83.0)
Final diagnosis, n (%)		
Benign	198 (76)	52 (47)**
Malignant	63 (24)	7 (6)††
Material insufficient for diagnosis by Endorette®	0	12 (11)
Material insufficient for diagnosis by D&C	0	5 (5)
Material insufficient for diagnosis by hysteroscopy	0	4 (4)
No diagnostic procedure performed	0	30 (27)
Final diagnosis obtained by, n (%)		
Hysteroscopy	162 (62)	22 (20)
Hysterectomy	85 (33)	7 (6)
D&C	14 (5)	16 (15)
Endorette® sampling	0	35 (32)
No diagnostic procedure performed	0	30 (27)

Continued

Table 1 Continued

	Included (n = 261)	Excluded (n = 110)
Histological diagnosis, n (%)		
Polyp	129 (49)	14 (13)
Atrophic endometrium	24 (9)	15 (14)
Submucous myoma	14 (5)	2 (2)
Hormone influenced endometrium	9 (3)	13 (12)
Polyp and submucous myoma	7 (3)	1 (1)
Hyperplasia		
without atypia	10 (4)	5 (5)
with atypia	5 (2)	2 (2)
Endometroid adenocarcinoma	55 (21)	7 (6)
Clear cell adenocarcinoma	2 (1)	0
Serous papillary adenocarcinoma	3 (1)	0
Carcinosarcoma	3 (1)	0
Stage of malignant tumors, n (% of all malignancies)		
I,	40 (63)	3 (43)
II	9 (14)	1 (14)
III	11 (18)	1 (14)
IV	1 (2)	–
Not staged	2 (3)	2 (29)

D&C, dilatation and curettage
* Information lacking in 10 patients
† Information lacking in 14 patients
‡ Information lacking in 7 patients
§ Information lacking in 5 patients
¶ Information lacking in 8 patients
** In 21 cases only Endorette® sampling
†† In 2 cases only Endorette® sampling

Table 2 Clinical background data and ultrasound results for women with benign and malignant endometrium in the study

	Benign (n = 198)	Malignant (n = 63)	p*
Age, (years); median (range)	65 (47 − 89)	74 (52 − 91)	0.0005
BMI, (kg/m²); median (range)†	28 (17 −50)	28 (19 − 42)	0.148
Nulliparous, n (%)‡	22 (11)	9 (14)	0.493
Hypertension, n (%)	84 (42)	27 (43)	0.952
Diabetes, n (%)	26 (13)	9 (14)	0.816
Hormone replacement therapy, n (%)	70 (35)	9 (14)	0.001
Local estrogen, n (%)	28 (14)	2 (3)	0.008
Low dose oral estriol, n (%)	7 (4)	5 (8)	0.170
Seq. or c.c. HRT	35 (18)	2 (5)	0.001
Use of tamoxifen, n (%)	12 (6)	1 (2)	0.112
Use of any anticoagulant, n (%)	46 (23)	17 (27)	0.548
Use of warfarin, n (%)	23 (12)	1 (2)	0.005
Use of levothyroxine, n (%)	24 (12)	8 (13)	0.903
Endometrial thickness, mm; median (range)	10.2 (4.5 − 75.6)	20.8 (6.4 − 59.0)	0.0005
Vascularity index, (%); median (range)§	13.6 (1.2- 95.7)	43.2 (2.4 − 93.5)	0.0005
VAS score; median (range)¶	13 (0 − 99)	65 (0 −100)	0.0005

*Univariate logistic regression analysis

† Information lacking in 10 patients

‡ Information lacking in 7 patients

§Percentage vascularized endometrial area as assessed objectively by computer analysis

¶Color content of the endometrial scan as assessed subjectively on a visual analogue scale (1-100)

96

Table 3 Results of univariate and multivariate logistic regression with regard to predicting endometrial malignancy in women with postmenopausal bleeding and sonographic endometrial thickness ≥4.5 mm

	Odds ratio		p-value
	point estimate	95% CI	
Univariate logistic regression			
Age	1.057	(1.027 – 1.088)	0.0005
HRT	0.305	(0.142 – 0.654)	0.001
Use of warfarin	0.123	(0.016 – 0.928)	0.005
ET	1.121	(1.078 – 1.167)	0.0005
Vascularity index	1.082	(1.059 – 1.105)	0.0005
VAS score	1.061	(1.046 – 1.076)	0.0005
Multivariate logistic regression			
Model with clinical variables			
Model 1 (age, use of warfarin, HRT)			
Age	1.061	(1.030 – 1.093)	0.0005
HRT	0.338	(0.154 – 0.744)	0.004
Use of warfarin	0.085	(0.011 – 0.662)	0.001
Model with clinical variables and ET			
Model 2 (age, use of warfarin, ET)			
Age	1.053	(1.019 – 1.087)	0.001
Use of warfarin	0.077	(0.008 – 0.716)	0.003
ET	1.110	(1.066 – 1.157)	0.0005

Continued

97

Table 3 continued

	Odds ratio		
	point estimate	(95% CI)	p-value

Model with clinical variables, ET, VAS score

Model 3 (age, HRT, ET, VAS score)

Age	1.045	(1.004 – 1.088)	0.031
Any type of HRT	0.305	(0.108 – 0.862)	0.017
ET	1.060	(1.018 – 1.103)	0.002
VAS score	1.058	(1.041 – 1.075)	0.0005

Model with clinical variables, ET, vascularity index

Model 4 (age, HRT, ET, vascularity index)

Age	1.072	(1.028 – 1.118)	0.001
HRT	0.264	(0.086 – 0.806)	0.012
ET	1.088	(1.044 – 1.133)	0.000
Vascularity index	1.092	(1.064 – 1.121)	0.0005

ET, endometrial thickness; HRT, hormonal replacement therapy; VAS, visual analogue scale;
Odds ratios have been calculated for endometrial thickness expressed in mm, age in years, and vascularity index in percent

Table 4 Diagnostic performance with regard to predicting endometrial malignancy in women with postmenopausal bleeding and endometrial thickness $\geq$ 4.5 mm

	AUC [95% CI]	Cut-off value*	Sensitivity, % [95% CI]	Specificity, % [95% CI]	LR+ [95% CI]	LR- [95% CI]	Specificity in % at 90.5% (57/63) sensitivity [95% CI]
Age	0.66 (0.59 – 0.74)	72 years	60 (38/63) [47-72]	69 (137/198) [62-76]	1.96 [1.4-2.8]	0.57 [0.4-0.8]	30 (59/198) [24-37]
No use of HRT	0.61 (0.53 – 0.68)	NA	86 (54/63) [75-93]	35 (70/198) [29-42]	1.33 [1.1-1.6]	0.40 [0.1-0.8]	NA
ET	0.80 (0.73 – 0.86)	15.3mm	70 (44/63) [57-81]	81 (161/198) [75-86]	3.74 [2.5-5.8]	0.37 [0.2-0.5]	45 (90/198) [38-53]
VI	0.82 (0.76 – 0.89)	29.4%	70 (44/63) [57-81]	88 (174/198) [83-92]	5.76 [3.6-10.1]	0.34 [0.2-0.5]	44 (88/198) [37-52]
VAS score	0.84 (0.77 – 0.90)	47.5	67 (42/63) [54-78]	95 (188/198) [91-98]	13.20 [6.6-34.4]	0.35 [0.2-0.5]	28 (55/198) [22-35]
No use of warfarin	0.55 (0.47 – 0.63)	NA	98 (62/63) [91-100]	12 (23/198) [8-17]	1.11 [1.0-1.2]	0.14 †	NA
Model 1 (age, use of warfarin, HRT)	0.74 (0.67 – 0.81)	0.17	89 (56/63) [78-95]	50 (99/198) [43-57]	1.78 [1.4-2.2]	0.22 [0.1-0.5]	46 (92/198) [39-54]

Continued

Table 4 continued

	AUC [95% CI]	Cut-off value*	Sensitivity, % [95% CI]	Specificity, % [95% CI]	LR+ [95% CI]	LR- [95% CI]	Specificity in % at 90.5% (57/63) sensitivity [95% CI]
Model 2 (age, use of warfarin, ET)	0.82 (0.76 – 0.87)	0.19	84 (53/63) [73-92]	66 (131/198) [59-73]	2.49 [1.9-3.3]	0.24 [0.1-0.4]	50 (99/198) [43-57]
Model 3 (age, HRT, ET, VAS score)	0.89 (0.84 – 0.94)	0.51	70 (44/63) [57-81]	96 (191/198) [93-99]	19.76 [9.0-66.1]	0.31 [0.2-0.5]	55 (108/198) [47-61]
Model 4 (age, HRT, ET, VI) ‡	0.91 (0.87 – 0.95)	0.14	89 (56/63) [78-95]	77 (152/198) [70-82]	3.83 [2.8-5.5]	0.14 [0.0-0.3]	71 (141/198) [64-77]

AUC, area under the receiver operating characteristic curve; CI, confidence interval; LR+, positive likelihood ratio; LR-, negative likelihood ratio; ET, endometrial thickness; HRT, any type of hormonal replacement therapy; VAS, visual analogue scale; VI, vascularity index; NA, not applicable
*Mathematically optimal cut-off value corresponding to the point on the receiver operating characteristic curve situated farthest from the reference line
† 4.839e-005 – 1.062
‡ The probability of malignancy for this model is calculated as $[e^z/(1+e^z)]$ where e = 2.718 (base value of natural logarithms) and z is calculated as follows: z = -9.356 + (0.084 x endometrial thickness in mm) + (0.088 x VI in percent) + (0.069 x age in years) + (-1.333 x use of HRT; coded 1 if used and 0 if not used). Using a risk cutoff of 0.11 to indicate malignancy, this model had sensitivity 90.5%, specificity 71%, LR+ 3.14 and LR- 0.13

100

Table 5 Intra-observer repeatability and reliability for analysis of vascularity index

Median of values (n = 40)	Range of values (n = 40)	ICC (95% CI)	Difference between two analyses by the same observer (n=20)	
			Mean (95% CI)	Repeatability coefficient (95% CI)
18.23	2.77 – 86.62	1.00 (1.00 - 1.00)	0.026 (-0.11 – 0.16)	0.56 (0.428 - 0.807)

ICC, intra-class correlation coefficient; CI, confidence interval

Table 6 Intra-observer repeatability and reliability and inter-observer reproducibility and reliability for analysis of visual analogue scale (VAS) score

	Median of values (n = 60)	Range of values (n = 60)	ICC (95% CI)	Difference between two analyses (n=30)		
				Mean (95% CI)	Repeatability coefficient (95% CI)	Limits of agreement
Intra-observer repeatability and reliability						
Observer 1	17.50	0 – 96	0.96 (0.89 – 0.98)	-4.03 (-6.98 - (- 1.09))	17.12 (13.7-22.9)	-
Observer 2	20.50	0 – 95	0.96 (0.88 – 0.98)	4.00 (1.48 – 6.46)	15.02 (12.0-20.1)	-
Inter-observer reproducibility and reliability						
	19.00	0 – 95	0.96 (0.92 – 0.98)	-1.90 (-4.77 – 0.97)	-	-17.0–13.2

ICC, intra-class correlation coefficient; CI, confidence interval

III

Three-dimensional ultrasound imaging for discrimination between benign and malignant endometrium in women with postmenopausal bleeding and sonographic endometrial thickness of at least 4.5 mm

G. OPOLSKIENE, P. SLADKEVICIUS, L. JOKUBKIENE and L. VALENTIN

Department of Obstetrics and Gynecology, Malmö University Hospital, Lund University, Malmö, Sweden

ABSTRACT

Objectives *To determine whether endometrial volume or power Doppler indices as measured by three-dimensional (3D) ultrasound imaging can discriminate between benign and malignant endometrium, to compare their diagnostic performance with that of endometrial thickness measurement using two-dimensional (2D) ultrasound examination, and to determine whether power Doppler indices add any diagnostic information to endometrial thickness or volume.*

Methods *Sixty-two patients with postmenopausal bleeding and endometrial thickness ≥ 4.5 mm underwent transvaginal 2D gray-scale and 3D power Doppler ultrasound examination of the corpus uteri. The endometrial volume was calculated, along with the vascularization index (VI), flow index and vascularization flow index (VFI) in the endometrium and in a 2-mm 'shell' surrounding the endometrium. The 'gold standard' was the histological diagnosis of the endometrium obtained by hysteroscopic resection of focal lesions, dilatation and curettage or hysterectomy. Receiver–operating characteristics (ROC) curves were drawn for all measurements to evaluate their ability to distinguish between benign and malignant endometrium. Multivariate logistic regression analysis was used to create mathematical models to estimate the risk of endometrial malignancy.*

Results *There were 49 benign and 13 malignant endometria. Endometrial thickness and volume were significantly larger in malignant than in benign endometria, and flow indices in the endometrium and endometrial shell were significantly higher. The area under the ROC curve (AUC) of endometrial thickness was 0.82, that of endometrial volume 0.78, and that of the two best power Doppler variables (VI and VFI in the endometrium) 0.82 and 0.82. The best logistic regression model for predicting malignancy contained the variables endometrial thickness (odds ratio 1.2; 95% CI, 1.04–1.30; P = 0.004) and VI in the endometrial 'shell' (odds ratio 1.1; 95% CI, 1.02–1.23; P = 0.01). Its AUC was 0.86. Using its mathematically optimal risk cut-off value (0.22), the model correctly classified seven more benign cases but two fewer malignant cases than the best endometrial thickness cut-off (11.8 mm). Models containing endometrial volume and flow indices performed less well than did endometrial thickness alone (AUC, 0.79 vs. 0.82).*

Conclusions *The diagnostic performance for discrimination between benign and malignant endometrium of 3D ultrasound imaging was not superior to that of endometrial thickness as measured by 2D ultrasound examination, and 3D power Doppler imaging added little to endometrial thickness or volume. Copyright © 2009 ISUOG. Published by John Wiley & Sons, Ltd.*

INTRODUCTION

Postmenopausal bleeding is an early clinical sign of endometrial malignancy. The finding of a thin (< 5 mm) endometrium at transvaginal ultrasound examination in women with postmenopausal bleeding rules out about 99% of endometrial cancers[1]. Therefore, endometrial sampling is usually considered necessary only in women with postmenopausal bleeding and endometrial thickness ≥ 5 mm[1,2]. However, many women with postmenopausal bleeding and endometrial thickness ≥ 5 mm do not have

ORIGINAL PAPER

endometrial cancer and some do not have any endometrial pathology at all, but will still undergo – perhaps unnecessarily – interventional diagnostic procedures such as dilatation and curettage (D&C) or hysteroscopy. In the high-risk group of women with postmenopausal bleeding and a thick endometrium, the thicker the endometrium the higher the risk of malignancy[3,4]. Diagnostic methods other than a simple measurement of endometrial thickness, e.g. Doppler examination of the endometrium, might be helpful for discrimination between benign and malignant endometrium in women with postmenopausal bleeding and thick endometrium, and so be used to select those women who might best benefit from an invasive diagnostic procedure[3,4].

Subjective evaluation of the color content of the endometrial scan or of the morphology of endometrial blood vessels using conventional two-dimensional (2D) power Doppler ultrasound imaging can discriminate between benign and malignant endometrium, and, when added to gray-scale ultrasound imaging, may improve discrimination in women with postmenopausal bleeding and endometrial thickness ≥ 4.5 mm[3,4]. However, these methods of assessing endometrial vascularity have the disadvantage of being purely subjective. The color content in a representative 2D power Doppler image of the endometrium can also be quantified objectively using dedicated software[3], but this method, too, includes an element of subjectivity, because the image selected for objective quantification is chosen subjectively. Moreover, a single power Doppler image is likely to reflect the vascularization of the endometrium less well than multiple power Doppler images in an endometrial volume obtained by three-dimensional (3D) power Doppler imaging. In theory, therefore, 3D power Doppler ultrasound examination might be superior to 2D power Doppler ultrasonography when estimating the risk of endometrial malignancy in a woman with postmenopausal bleeding. Some have claimed that endometrial volume measurements taken using 3D ultrasound imaging discriminate better between benign and malignant endometrium than do endometrial thickness measurements[5,6].

The aim of this study was to determine whether endometrial volume or power Doppler indices as measured by 3D ultrasound imaging can discriminate between benign and malignant endometrium in women with postmenopausal bleeding and endometrial thickness ≥ 4.5 mm (rounded up to 5 mm), to compare the diagnostic performance of the 3D measurements with that of endometrial thickness measurements obtained by 2D ultrasonography, and to determine whether 3D power Doppler indices add any diagnostic information to endometrial thickness or volume.

METHODS

The Ethics Committee of Lund University approved the study protocol. Informed consent was obtained from all participants after the nature of the procedures had been fully explained.

Patients with postmenopausal bleeding examined in our ultrasound unit and found to have endometrial thickness ≥ 4.5 mm (rounded up to 5 mm) were potentially eligible for inclusion in our study, i.e. for transvaginal 3D ultrasound examination. A woman was considered to be postmenopausal if she reported absence of menstruation for at least 1 year after the age of 40 years provided that the amenorrhea was not explained by pregnancy, medication or disease. Postmenopausal bleeding was defined as any vaginal bleeding in a postmenopausal woman not on hormone replacement therapy, or as unscheduled vaginal bleeding in a postmenopausal woman receiving hormone replacement therapy.

In all, 151 consecutive patients with postmenopausal bleeding were examined in our postmenopausal bleeding clinic by one of two coauthors of this paper (L.V. or P.S.), who are both experienced ultrasound examiners. Of these, 95 had an endometrial thickness ≥ 4.5 mm as measured by ultrasound examination. Twenty of the 95 women did not undergo a 3D examination because the 3D ultrasound system was not available.

Seventy-five women underwent both 2D and 3D transvaginal ultrasound examination by L.V. or P.S. following a standardized research protocol. A GE Voluson 730 Expert ultrasound system equipped with a 6–9-MHz transvaginal transducer (GE Healthcare, Zipf, Austria) was used. Endometrial thickness was measured on a sagittal section through the uterus using 2D gray-scale ultrasound imaging as described previously[7]. Then the system was switched into the power Doppler mode. Identical preinstalled settings were used for all patients: frequency 'normal', pulse repetition frequency 0.6 kHz, gain 4.0, wall motion filter 'low' (40 Hz). Finally, the 3D mode was activated, the corpus uteri (without the cervix) was centralized within the 3D sector on the screen so that it filled the whole 3D sector, and a volume of the corpus uteri was acquired using the setting 'Quality high 2' for the 3D sweep speed. The patient was asked to remain still during the acquisition. The multiplanar display resulting from the acquisition was examined to ensure that a complete volume of the uterine corpus had been captured. The volumes were saved for later analysis. When more than one volume of the corpus uteri had been acquired from the same patient, only the best one was stored electronically for later analysis. Volumes with color artifacts were not stored.

The stored 3D volumes of the corpus uteri were analyzed by one observer (L.J.). Endometrial volume, and three power Doppler indices in the endometrium (Figure 1) and in a 2-mm 'shell' (i.e. subendometrium) surrounding the endometrium (Figure 2), were calculated using VOCAL™ software (GE Healthcare). Rotation steps of 30° were used. The A-plane (sagittal view of the uterus) was rotated around the *y*-axis with all measurements being conducted on the A-plane. The flow indices calculated were vascularization index (VI), flow index (FI) and vascularization flow index (VFI). These indices have
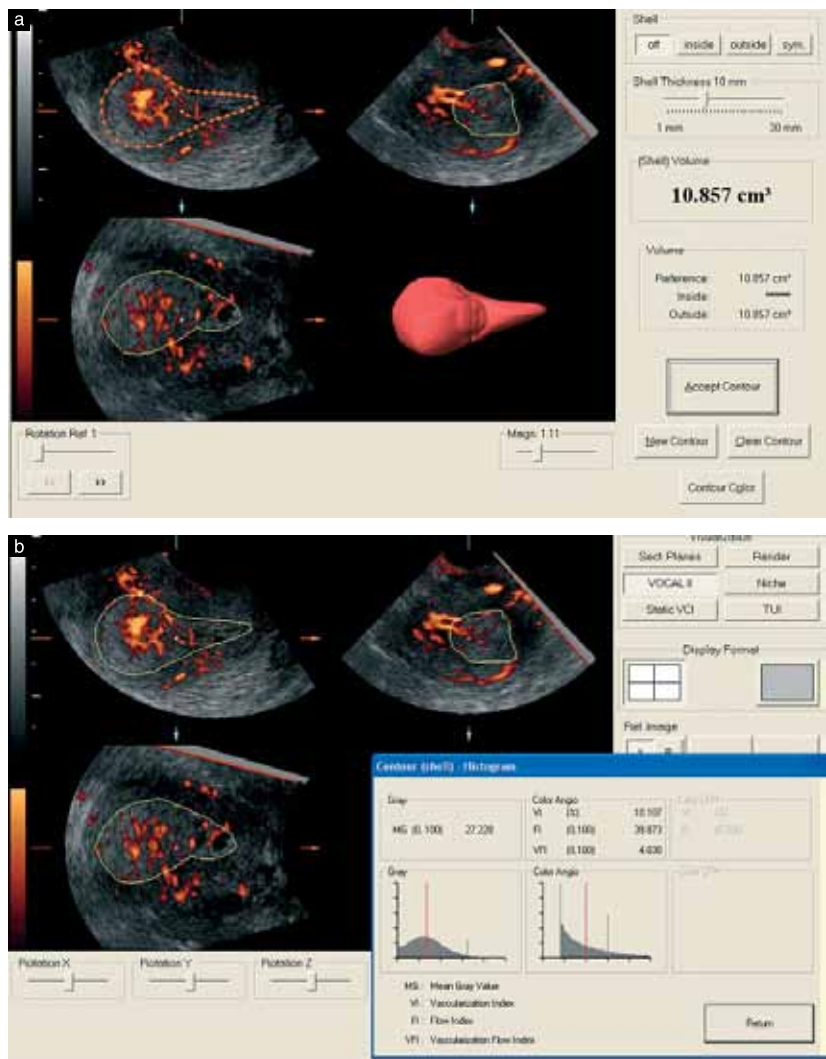
Figure 1 Calculation using VOCAL™ software of endometrial volume (a) and power Doppler indices (vascularization index (VI), flow index (FI) and vascularization flow index (VFI)) in the endometrium (b).

been described in several publications, such as Jokubkiene et al.[8].

The histological diagnosis of the endometrium, obtained by hysteroscopic resection of focal lesions, D&C or hysterectomy, was our 'gold standard'. Patients with a diagnosis based only on an Endorette® sample (Medscand AB, Malmö, Sweden) were excluded, because

we wanted to be sure that only representative endometrial samples were used to establish the final diagnosis[9]. Staging of endometrial malignancies was done by the attending physician in accordance with the classification system recommended by the International Federation of Gynecology and Obstetrics[10].
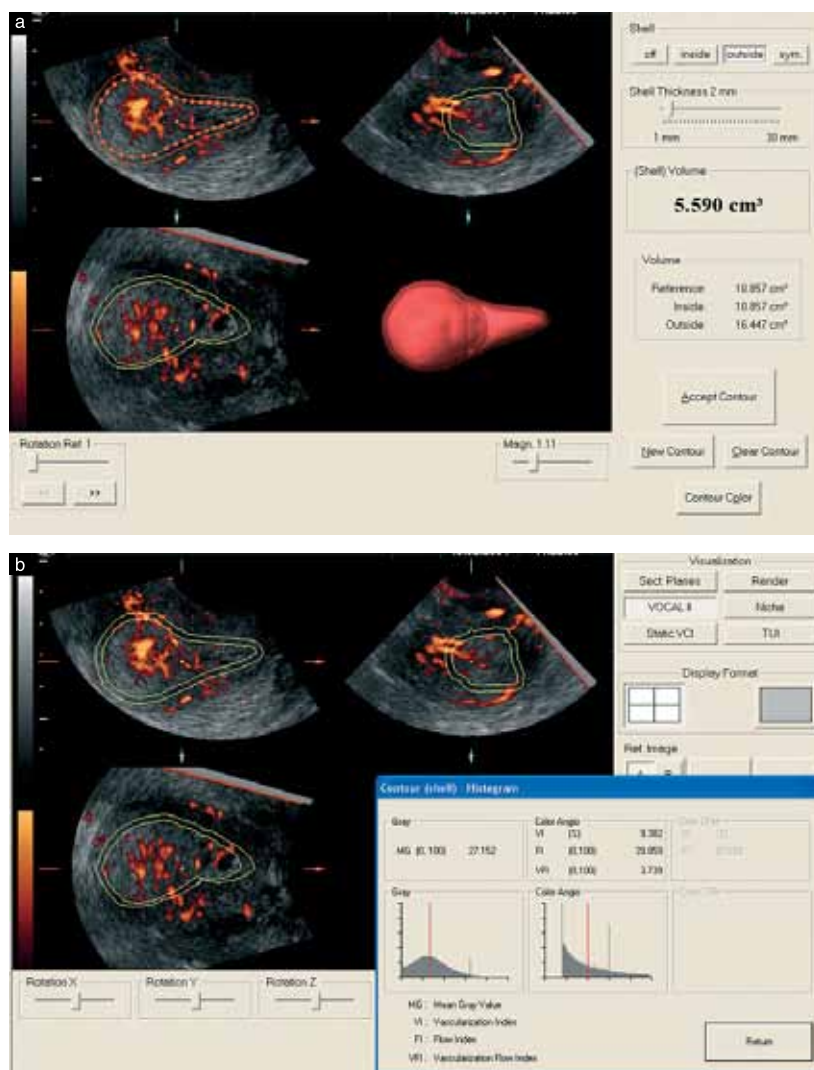
Figure 2 Calculation using VOCAL™ software of subendometrial volume, i.e. the 2-mm 'shell' surrounding the endometrium (a) and power Doppler indices (vascularization index (VI), flow index (FI) and vascularization flow index (VFI)) in the subendometrium (b).

Statistical analysis

Statistical calculations were performed using SPSS versions 12.2 and 16.0 (SPSS Inc., Chicago, IL, USA). The 95% confidence intervals (CI) for the likelihood ratios were calculated using StatXact software, version 4 (Cytel Inc., Cambridge, MA, USA). A possible relationship between single ultrasound variables and endometrial malignancy was determined using univariate logistic regression with the likelihood ratio test, two-tailed $P < 0.05$ being considered statistically significant. Multivariate logistic regression analysis with the likelihood ratio test was used to create mathematical models to calculate the risk of endometrial malignancy. To avoid

overfitting, only two variables were allowed in the models, i.e. endometrial thickness or endometrial volume and one additional ultrasound variable. Receiver–operating characteristics (ROC) curves were drawn for all ultrasound measurements and for all logistic regression models to evaluate their ability to distinguish between benign and malignant endometrium. The area under the ROC curve (AUC) and the 95% CI of this area were calculated. If the lower limit of the CI for the area under the ROC curve was > 0.5, the variable was considered to have discriminatory potential. The sensitivity, specificity, positive and negative likelihood ratios (LR+ and LR−, respectively) of the mathematically best cut-off value for predicting endometrial malignancy were calculated for the ultrasound variables and for the models. The mathematically best cut-off value for predicting endometrial malignancy was defined as that corresponding to the point on the ROC curve situated farthest from the reference line. We defined the best diagnostic test as the one with the largest area under the ROC curve. We also calculated the specificity corresponding to a sensitivity of 85% for all ultrasound variables and models, because our results showed that the sensitivity obtained using the mathematically best cut-off value for endometrial thickness to predict malignancy was 85%.

Intraobserver repeatability and reliability[11] of measurements of endometrial and subendometrial volume and flow indices were determined for the member of our team (L.J.) who performed all the volume analyses for this study. To determine the intraobserver repeatability and reliability we used the volumes of 30 patients in our study sample. These patients were selected from our statistical data sheet so as to include 10 patients with malignant endometrium and 20 with benign endometrium. The selection was done to cover the whole study period but without looking at the results of the ultrasound examinations. The 30 volumes were analyzed by L.J. twice, with approximately 1 week elapsing between the first and second analysis. Intraobserver repeatability was estimated as the difference between the two measurement results. The differences between the measured values were plotted against the mean of the two measurements (Bland–Altman plot) to assess the relationship between the differences and the magnitude of the measurements[12]. Systematic bias between the first and second analysis was estimated by calculating the 95% CI of the mean difference (mean difference ±2 standard error of the mean). If zero lay within this interval no bias was assumed to exist between the first and second measurement. Intraobserver repeatability was expressed as the repeatability coefficient. The absolute difference between two measurements on a subject is expected to differ by no more than the repeatability coefficient on 95% of occasions[11]. Intraobserver reliability was expressed as the intraclass correlation coefficient (ICC)[11]. The ICC was calculated using ANOVA (two-way random model–absolute agreement)[11]. The ICC indicates the proportion of the total variance in measurement results that can be explained by differences between the individuals examined. A high ICC indicates that the

measurements can be used to discriminate between individuals. The more variable the population investigated, the greater the ICC and the less variable the population, the smaller the ICC[11].

RESULTS

Thirteen of the 75 patients examined with both 2D and 3D gray-scale and power Doppler ultrasound imaging were excluded; seven women did not undergo any endometrial sampling at all, in four women only an endometrial biopsy using the Endorette® device was performed, in one case a large myoma obscured the endometrium so that a reliable analysis of the acquired volume was not possible, and one case was excluded because of technical problems. Of the 62 women finally included, 49 (79%) had benign endometrium and 13 (21%) had malignant endometrium (Table 1). Median age was 65 years, and 24 (39%) patients were on hormone replacement therapy (Table 1). Patients with malignant endometrium tended to be older (median age, 73 years vs. 63 years; $P = 0.085$), had a higher body mass index (median, 31 kg/m^2 vs. 27 kg/m^2; $P = 0.068$) and thicker endometrium (median, 16.2 mm vs. 9.6 mm; $P = 0.001$) than those with benign endometrium, and fewer used hormone replacement therapy (23% vs. 43%; $P = 0.181$).

Endometrial volume was larger and flow indices were higher both in the endometrium and in the endometrial shell in patients with malignant endometrium than in those with benign endometrium, but there was substantial overlap between the two categories (Table 2). The diagnostic performance of the various ultrasound

Table 1 Characteristics of the 62 patients included in the study

Characteristic	Median (range) or n (%)
Age (years)	65 (50–90)
Body mass index (kg/m^2)	27 (20–43)
Hormone replacement therapy	24 (39)
Systemic hormonal therapy	12 (19)
Local hormonal therapy	12 (19)
Tamoxifen therapy	3 (5)
Method used to obtain final diagnosis	
Hysteroscopic resection	38 (61)
Hysterectomy	15 (24)
D&C	9 (15)
Histological diagnosis	
Polyp	28 (45)
Atrophic endometrium	9 (15)
Myoma	4 (6)
Hyperplasia without atypia	4 (6)
Polyp and myoma	2 (3)
Estrogen influenced endometrium	2 (3)
Adenocarcinoma	12 (19)
Carcinosarcoma	1 (2)
Stage of malignant tumors ($n = 13$)	
I	9 (15)
II	2 (3)
III	1 (2)
Not staged	1 (2)

D&C, dilatation and curettage.

Table 2 Ultrasound results for patients with benign and malignant endometrium

Ultrasound result	Benign (n = 49)	Malignant (n = 13)	P‡
Endometrial thickness (mm)	9.6 (4.5–32.9)	16.2 (8.2–31.0)	0.001
Endometrial volume (cm³)	2.3 (0.52–36.93)	10.2 (1.36–29.79)	0.003
VI in the endometrium (%)	1.2 (0.00–66.54†)	11.7 (0.44–32.65)	0.008
FI in the endometrium*	24.3 (0.00–50.53)	34.5 (21.86–39.23)	0.001
VFI in the endometrium*	0.3 (0.00–33.62)	3.6 (0.10–12.47)	0.034
Endometrial shell volume (cm³)	2.5 (1.31–12.19)	5.7 (1.74–10.62)	0.001
VI in the shell (%)	1.4 (0.00–25.98)	4.2 (0.05–31.43)	0.002
FI in the shell*	25.4 (0.00–44.97)	34.5 (19.39–40.07)	0.013
VFI in the shell*	0.5 (0.00–11.22)	1.3 (0.01–12.60)	0.002

Data are median (range). *Flow index (FI) and vascularization flow index (VFI) can have any value from 1 to 100. †Ultrasound images of the case (atypical polypoid adenomyoma) with a vascularization index (VI) in the endometrium of 66.54% are shown in Figure 3. ‡Univariate logistic regression analysis.

variables is shown in Table 3. The best variables for discrimination between benign and malignant endometrium, i.e. endometrial thickness and endometrial VI and VFI, had similar diagnostic performance, all having an AUC of 0.82. Using the mathematically best cut-off value for endometrial thickness to predict malignancy (11.8 mm) the sensitivity was 85%, the specificity 71%, the LR+ 3.0 and the LR− 0.2. Using the mathematically best cut-off value for VI and VFI in the endometrium, the sensitivity was lower than that of endometrial thickness (69% and 69% vs. 85%), the specificity was higher (84% and 82% vs. 71%), and both the LR+ and LR− were higher.

The following variables added information to endometrial thickness when calculating the risk of malignancy: FI in the endometrium, VI in the shell of the endometrium, and VFI in the shell of the endometrium. The best logistic regression models for predicting malignancy (i.e. the models with the largest AUC) contained endometrial thickness and VFI in the endometrial shell (AUC 0.86) or endometrial thickness and VI in the endometrial shell (AUC 0.86) (Table 3). Because the model including VI manifested higher specificity at sensitivities ≥ 85% we decided to regard it as the best logistic regression model for discrimination between benign and malignant endometrium. The probability of malignancy using this model was calculated as $[e^z/(1+e^z)]$ where e = 2.718 (mathematical constant and base value of natural logarithms) and z = −4.076 + (0.152 × endometrial thickness in mm) + (0.115 × VI in the endometrial shell expressed as a percentage); the odds ratio for endometrial thickness was 1.2 (95% CI, 1.04–1.30; $P = 0.004$) and the odds ratio for VI in the endometrial shell was 1.1 (95% CI, 1.02–1.23; $P = 0.01$). The mathematically optimal risk cut-off value of this model (0.22) had a sensitivity of 69%, a specificity of 86%, a LR+ of 4.8 and a LR− of 0.4 (Table 3). Using the mathematically best cut-off value of the model, instead of using that of endometrial thickness for discrimination between benign and malignant endometrium, seven additional benign cases were correctly classified but two additional endometrial malignancies were missed.

Table 3 Diagnostic performance of single ultrasound variables and logistic regression models

	Area under ROC curve (95% CI)	Cut-off value*	Sensitivity (% (n))	Specificity (% (n))	LR+ (95% CI)	LR− (95% CI)
Single ultrasound variables						
Endometrial thickness	0.82 (0.69–0.94)	11.8 mm	85 (11/13)	71 (35/49)	3.0 (1.3–7.1)	0.2 (0.0–0.7)
Endometrial volume	0.78 (0.61–0.94)	5.3 cm³	69 (9/13)	88 (43/49)	5.7 (1.7–28.6)	0.4 (0.0–0.8)
VI in the endometrium	0.82 (0.69–0.94)	5.0%	69 (9/13)	84 (41/49)	4.2 (1.4–16.0)	0.4 (0.0–0.8)
FI in the endometrium	0.81 (0.68–0.94)	30.2	77 (10/13)	82 (40/49)	4.2 (1.5–13.7)	0.3 (0.0–0.7)
VFI in the endometrium	0.82 (0.69–0.94)	1.5	69 (9/13)	82 (40/49)	3.8 (1.2–12.2)	0.4 (0.0–0.8)
Endometrial shell volume	0.77 (0.61–0.94)	4.1 cm³	69 (9/13)	88 (43/49)	5.7 (1.7–28.6)	0.4 (0.0–0.8)
VI in the shell	0.75 (0.60–0.90)	3.4%	69 (9/13)	73 (36/49)	2.6 (1.0–7.1)	0.4 (0.1–1.0)
FI in the shell	0.72 (0.57–0.87)	32.0	62 (8/13)	82 (40/49)	3.3 (1.0–12.2)	0.5 (0.1–0.9)
VFI in the shell	0.75 (0.60–0.90)	1.0	69 (9/13)	73 (36/49)	2.6 (1.0–7.1)	0.4 (0.1–1.0)
Logistic regression models						
ET and VI in the shell	0.86 (0.75–0.96)	0.22	69 (9/13)	86 (42/49)	4.8 (1.5–20.4)	0.4 (0.0–0.8)
ET and VFI in the shell	0.86 (0.75–0.96)	0.17	77 (10/13)	80 (39/49)	3.8 (1.3–10.9)	0.3 (0.0–0.8)
ET and FI in the endometrium	0.84 (0.72–0.95)	0.31	69 (9/13)	88 (43/49)	5.7 (1.7–28.6)	0.4 (0.0–0.8)
EV and VI in the shell	0.79 (0.63–0.95)	0.22	69 (9/13)	88 (43/49)	5.7 (1.7–28.6)	0.4 (0.0–0.8)
EV and VFI in the shell	0.79 (0.63–0.95)	0.24	69 (9/13)	94 (46/49)	11.3 (2.5–164)	0.3 (0.0–0.7)

*Mathematically optimal cut-off value as defined in the text. ET, endometrial thickness; EV, endometrial volume; FI, flow index; LR−, negative likelihood ratio; LR+, positive likelihood ratio; ROC, receiver–operating characteristics; VFI, vascularization flow index; VI, vascularization index.
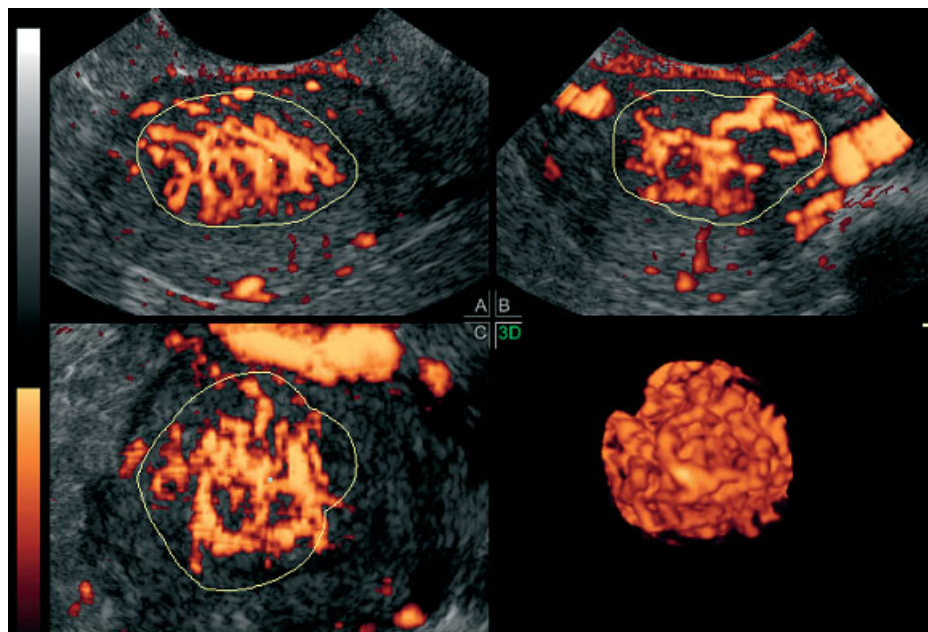
Figure 3 Three-dimensional power Doppler image of atypical polypoid adenomyoma. Sectional planes through the uterus (A, B and C) and a three-dimensional rendered image (3D) of the vessels in the endometrium are shown. In this case, the endometrium was very richly vascularized (VI, 66.54%).
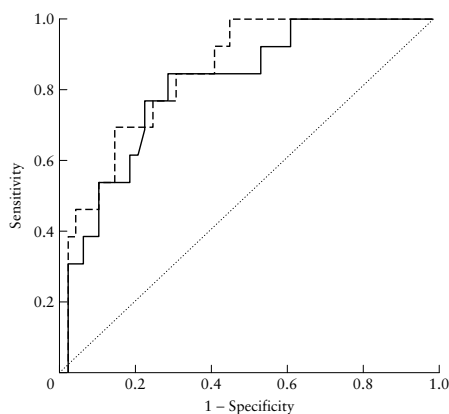

Figure 4 Receiver–operating characteristics curves for endometrial thickness (⸻) and for the best logistic regression model containing two variables (endometrial thickness and vascularization index in the shell, - - - -) in the discrimination between benign and malignant endometrium. The diagonal line is the reference line.

The ROC curves for endometrial thickness and the best logistic regression model are shown in Figure 4.

At a sensitivity of 85%, only the model containing endometrial thickness and VFI in the shell, or endometrial thickness and VI in the shell had a specificity as high as (71%) or similar (69%) to that of endometrial thickness. At sensitivity 85%, the specificity of VI and VFI in the endometrium was 61%. Models containing endometrial volume and flow indices performed less well than models containing endometrial thickness and flow indices. They also performed less well than endometrial thickness alone (Table 3).

Results with regard to intraobserver repeatability and reliability are shown in Table 4. There was no systematic difference between the first and second measurement by the same observer, and there was no clear relationship between the difference between the two measurements and the mean of the two measurements when we analyzed Bland–Altman plots. Intraobserver reliability was very high for all ultrasound variables (all ICC values > 0.95).

DISCUSSION

Our results showed that, even though the endometrial volumes were significantly larger and flow indices significantly higher in malignant than in benign endometria, neither volume measurements nor flow indices discriminated

Table 4 Intraobserver repeatability and reliability for analysis of volume and vascular indices of the endometrium and endometrial shell in 30 patients

Parameter	Median (range) (n = 60)	ICC (95% CI)	Difference between two analyses by the same observer (n = 30)	
			Mean (95% CI)	Repeatability coefficient
Endometrial volume (cm³)	4.81 (1.03–45.80)	1.000 (0.999–1.000)	−0.051 (−0.143 to 0.041)	0.485
VI in the endometrium (%)	3.24 (0.00 −29.39)	0.993 (0.985–0.997)	0.001 (−0.400 to 0.403)	2.071
FI in the endometrium	32.21 (0.00–42.42)	0.998 (0.976–0.994)	−0.090 (−0.615 to 0.434)	2.713
VFI in the endometrium	1.20 (0.00–12.16)	0.988 (0.975–0.994)	−0.042 (−0.248 to 0.163)	1.065
Endometrial shell volume (cm³)	3.49 (1.71–13.76)	0.997 (0.994–0.999)	0.020 (−0.065 to 0.107)	0.447
VI in the shell (%)	2.32 (0.00–31.16)	0.978 (0.954–0.989)	0.152 (−0.440 to 0.743)	3.066
FI in the shell	31.96 (18.95–50.29)	0.950 (0.899–0.976)	−0.390 (−1.273 to 0.492)	4.618
VFI in the shell	0.71 (0.00–13.31)	0.974 (0.946–0.988)	0.034 (−0.227 to 0.296)	1.351

FI, flow index; ICC, intraclass correlation coefficient; VFI, vascularization flow index; VI, vascularization index.

well between benign and malignant endometria, and none of the 3D measurements was clearly superior to simple endometrial thickness measurements taken using 2D ultrasound examination. They also showed that 3D flow indices added little to endometrial thickness, the effect of adding flow index (VI in subendometrium) to endometrial thickness being an increase in specificity at high sensitivities.

To the best of our knowledge, only three other studies have tried to determine the diagnostic performance of 3D ultrasound imaging with regard to discrimination between benign and malignant endometrium[5,6,13]. A fourth study examined the ability of 3D ultrasound examination to discriminate between endometrial carcinoma and hyperplasia[14], and a fifth study the ability to distinguish cancer and hyperplasia from other benign conditions[15]. All five studies reported on the diagnostic performance of endometrial volume measurements[5,6,13–15] and three of them also on that of 3D power Doppler flow indices[13–15]. In all five studies endometrial volume measurements were superior to endometrial thickness measurements, endometrial volume manifesting excellent diagnostic properties in one of the studies[5], good diagnostic performance in another two studies[6,13] but very limited diagnostic performance in the remaining two studies[14,15]. In our study, the diagnostic performance of endometrial volume was not superior to that of endometrial thickness. In the study by Odeh *et al.*[15] endometrial volume was superior to the flow indices for discrimination between hyperplasia/malignant endometrium and benign endometrium other than hyperplasia, whereas in the investigations by Mercé *et al.*[14] and Alcazar *et al.*[13] the flow indices were superior to endometrial volume for discrimination between endometrial carcinoma and endometrial hyperplasia, and between benign and malignant endometrium, respectively. However, the diagnostic performance of the flow indices was rather poor, which is in agreement with the results of our study.

The differences in results between the five published studies as well as the differences in results between our study and the other five studies can almost certainly be explained by substantial differences in study populations and study design. Even though all the publications cited include only women with abnormal vaginal bleeding, there are differences in menopausal status, rate of endometrial cancer, mix of benign histologies, use of hormone replacement therapy and endometrial thickness. In addition to there being differences in group comparisons between the studies (cancer vs. hyperplasia[14], cancer or hyperplasia vs. other benign histology[15], benign vs. malignant[5,6,13]), there are also differences in the methods used to determine diagnostic performance. In only one of the published studies[6] was there a definition of 'best cut-off' for predicting malignancy.

We know of no other published study that has explored the effect of adding information on 3D flow indices to endometrial thickness on the diagnostic performance with regard to discriminating between benign and malignant endometrium. In our study, the effect was rather small. This is in agreement with adding information on 2D power Doppler imaging to endometrial thickness having only a small positive effect on diagnostic performance[3,4]. It seems that endometrial thickness is a very powerful predictor of endometrial malignancy even in a high-risk group of women with postmenopausal bleeding and endometrial thickness ≥ 4.5 mm, the risk of endometrial cancer increasing with increasing endometrial thickness[3,4].

We have found five studies that examined the reproducibility of 3D measurements of endometrial volume[16–20] and four studies that examined the reproducibility of 3D flow indices[17–20]. In three studies the intraobserver reproducibility was assessed[16,17,20] and in four studies the interobserver reproducibility was estimated[16–19]. Only one study comprised exclusively postmenopausal patients, and in that study only the reproducibility of endometrial volume measurements – not that of flow indices – was determined[16]. Despite the study populations and study designs being different, all studies reported high ICC values (≥ 0.80), i.e. in all studies the measurements discriminated well

between individuals[11,21]. However, only three studies reported on the magnitude of the differences (limits of agreement)[18,19,20]; these were large in one study[19], small in one[18] and difficult to interpret in another[20]. In our study, the repeatability coefficient seemed to be quite large in comparison with the median VI and VFI values, but small in comparison with the median volume and FI values. Our reproducibility results are in agreement with previously published results: the ICC values were high for all measured variables. That the ICC value is high despite the repeatability coefficient for VI and VFI being quite large in comparison with the average VI and VFI values is likely to be explained by the substantial variability in our measurement results[11,21]. For the purpose of this study, the reliability of our measurements was clearly acceptable.

To sum up, we found that the diagnostic performance of 3D ultrasound imaging with regard to discriminating between benign and malignant endometrium was not superior to that of endometrial thickness measured by 2D ultrasound examination, and that 3D power Doppler flow indices added little to endometrial thickness or volume.

ACKNOWLEDGMENTS

REFERENCES

1. Smith-Bindman R, Kerlikowske K, Feldstein VA, Subak L, Scheidler J, Segal M, Brand R, Grady D. Endovaginal ultrasound to exclude endometrial cancer and other endometrial abnormalities. *JAMA* 1998; **280**: 1510–1517.
2. Gull B, Karlsson B, Milsom I, Granberg S. Can ultrasound replace dilation and curettage? A longitudinal evaluation of postmenopausal bleeding and transvaginal sonographic measurement of the endometrium as predictors of endometrial cancer. *Am J Obstet Gynecol* 2003; **188**: 401–408.
3. Epstein E, Skoog L, Isberg PE, De Smet F, De Moor B, Olofsson PA, Gudmundsson S, Valentin L. An algorithm including results of gray-scale and power Doppler ultrasound examination to predict endometrial malignancy in women with postmenopausal bleeding. *Ultrasound Obstet Gynecol* 2002; **20**: 370–376.
4. Opolskiene G, Sladkevicius P, Valentin L. Ultrasound assessment of endometrial morphology and vascularity to predict endometrial malignancy in women with postmenopausal bleeding and sonographic endometrial thickness > or = 4.5 mm. *Ultrasound Obstet Gynecol* 2007; **30**: 332–340.
5. Gruboeck K, Jurkovic D, Lawton F, Savvas M, Tailor A, Campbell S. The diagnostic value of endometrial thickness and volume measurements by three-dimensional ultrasound in patients with postmenopausal bleeding. *Ultrasound Obstet Gynecol* 1996; **8**: 272–276.
6. Yaman C, Habelsberger A, Tews G, Polz W, Ebner T. The role of three-dimensional volume measurement in diagnosing endometrial cancer in patients with postmenopausal bleeding. *Gynecol Oncol* 2008; **110**: 390–395.
7. Sladkevicius P, Valentin L, Marsal K. Endometrial thickness and Doppler velocimetry of the uterine arteries as discriminators of endometrial status in women with postmenopausal bleeding: a comparative study. *Am J Obstet Gynecol* 1994; **171**: 722–728.
8. Jokubkiene L, Sladkevicius P, Valentin L. Does three-dimensional power Doppler ultrasound help in discrimination between benign and malignant ovarian masses? *Ultrasound Obstet Gynecol* 2007; **29**: 215–225.
9. Epstein E, Skoog L, Valentin L. Comparison of Endorette and dilatation and curettage for sampling of the endometrium in women with postmenopausal bleeding. *Acta Obstet Gynecol Scand* 2001; **80**: 959–964.
10. Shepherd JH. Revised FIGO staging for gynaecological cancer. *Br J Obstet Gynaecol* 1989; **96**: 889–892.
11. Bartlett JW, Frost C. Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. *Ultrasound Obstet Gynecol* 2008; **31**: 466–475.
12. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; **1**: 307–310.
13. Alcazar JL, Galvan R. Three-dimensional power Doppler ultrasound scanning for the prediction of endometrial cancer in women with postmenopausal bleeding and thickened endometrium. *Am J Obstet Gynecol* 2009; **200**: 44.e1–44.e6.
14. Mercé LT, Alcazar JL, Lopez C, Iglesias E, Bau S, Alvarez de los Heros J, Bajo JM. Clinical usefulness of 3-dimensional sonography and power Doppler angiography for diagnosis of endometrial carcinoma. *J Ultrasound Med* 2007; **26**: 1279–1287.
15. Odeh M, Vainerovsky I, Grinin V, Kais M, Ophir E, Bornstein J. Three-dimensional endometrial volume and 3-dimensional power Doppler analysis in predicting endometrial carcinoma and hyperplasia. *Gynecol Oncol* 2007; **106**: 348–353.
16. Yaman C, Ebner T, Jesacher K, Obermayr G, Polz W, Tews G. Reproducibility of three-dimensional ultrasound endometrial volume measurements in patients with postmenopausal bleeding. *Ultrasound Obstet Gynecol* 2002; **19**: 282–286.
17. Raine-Fenning NJ, Campbell BK, Clewes JS, Kendall NR, Johnson IR. The reliability of virtual organ computer-aided analysis (VOCAL) for the semiquantification of ovarian, endometrial and subendometrial perfusion. *Ultrasound Obstet Gynecol* 2003; **22**: 633–639.
18. Raine-Fenning NJ, Campbell BK, Clewes JS, Kendall NR, Johnson IR. The interobserver reliability of three-dimensional power Doppler data acquisition within the female pelvis. *Ultrasound Obstet Gynecol* 2004; **23**: 501–508.
19. Alcazar JL, Merce LT, Manero MG, Bau S, Lopez-Garcia G. Endometrial volume and vascularity measurements by transvaginal 3-dimensional ultrasonography and power Doppler angiography in stimulated and tumoral endometria: an interobserver reproducibility study. *J Ultrasound Med* 2005; **24**: 1091–1098.
20. Merce LT, Alcazar JL, Engels V, Troyano J, Bau S, Bajo JM. Endometrial volume and vascularity measurements by transvaginal three-dimensional ultrasonography and power Doppler angiography in stimulated and tumoral endometria: intraobserver reproducibility. *Gynecol Oncol* 2006; **100**: 544–550.
21. Burdock EI, Fleiss JL, Hardesty AS. A new view of interobserver agreement. *Pers Psychol* 1963; **16**: 373–384.

IV

Two- and three-dimensional saline contrast sonohysterography: interobserver agreement, agreement with hysteroscopy and diagnosis of endometrial malignancy

G. OPOLSKIENE, P. SLADKEVICIUS and L. VALENTIN

Department of Obstetrics and Gynecology, Malmö University Hospital, Lund University, Malmö, Sweden

ABSTRACT

Objectives *The aims of our study were to compare the interobserver reproducibility of two-dimensional (2D) and three-dimensional (3D) saline contrast sonohysterography (SCSH) and agreement of these techniques with hysteroscopy, and to determine which SCSH findings best discriminate between benign and malignant endometrium.*

Methods *Consecutive women with postmenopausal bleeding and endometrial thickness ≥ 4.5 mm underwent 2D and 3D SCSH; the results were videotaped and stored electronically, respectively, for analysis by two independent experienced examiners who were blinded to each other's results. A histological diagnosis was obtained by dilatation and curettage, hysteroscopic resection or hysterectomy. The hysteroscopist was blinded to the ultrasound results and used the same standardized research protocol to describe the uterine cavity as the ultrasound examiners.*

Results *Of 170 consecutive women with postmenopausal bleeding and endometrial thickness ≥ 4.5 mm, 84 (14 with endometrial malignancy) fulfilled our inclusion criteria. Hysteroscopy findings in 54 women (one with endometrial malignancy) were used to determine agreement with SCSH. Interobserver agreement of 2D and 3D SCSH was 95% (80/84) vs. 89% (75/84) with regard to presence of focal lesions, 89% (75/84) vs. 88% (74/84) for presence of focal lesions with irregular surface, 67% (54/81) vs. 63% (51/81) for number of focal lesions, and 77% (46/60) vs. 70% (42/60) for location of focal lesions. The agreement between 2D and 3D SCSH and hysteroscopy was 94% (51/54) vs. 93% (50/54) with regard to presence of focal lesions, 74% (40/54) vs. 76% (41/54) for presence of focal lesions with irregular surface, 63% (34/54) vs. 54% (29/54) for number of focal lesions, and 66%*

(29/44) vs. 64% (28/44) for location of focal lesions. The SCSH finding that best discriminated between benign and malignant endometrium was the presence of focal lesion(s) with irregular surface (for 2D SCSH: sensitivity 71%, specificity 97%, positive likelihood ratio 25, negative likelihood ratio 0.3; for 3D SCSH: sensitivity 43%, specificity 97%, positive likelihood ratio 15, negative likelihood ratio 0.6).

Conclusions *3D SCSH does not seem to be superior to 2D SCSH when performed by experienced ultrasound examiners either with regard to reproducibility, agreement with hysteroscopy findings or diagnosis of endometrial malignancy. The presence of focal lesion(s) with irregular surface is the best SCSH variable for discrimination between benign and malignant endometrium. Copyright © 2009 ISUOG. Published by John Wiley & Sons, Ltd.*

INTRODUCTION

Sonographic measurement of endometrial thickness can discriminate between women at low and high risk of endometrial cancer, women with endometrial thickness ≤ 4 mm being at low risk and those with endometrial thickness ≥ 5 mm being at high risk[1]. In women with a thick endometrium, the estimation of risk of malignancy can be refined by performing saline contrast sonohysterography (SCSH). The presence of focal lesions increases the odds of malignancy and the absence of focal lesions decreases them[2]. Irregular surface of a focal lesion might be associated with an even higher risk of malignancy[3].

SCSH can be performed using two-dimensional (2D) or three-dimensional (3D) ultrasound imaging, and some have suggested that 3D SCSH might be superior to 2D

ORIGINAL PAPER

SCSH[4,5]. The clinical usefulness of a diagnostic test depends not only on its diagnostic performance but also on its reproducibility. A few published studies have described the interobserver reproducibility of 2D and 3D SCSH, but they include mainly premenopausal women[6,7].

The aims of our study were to compare the interobserver reproducibility of 2D and 3D SCSH and agreement of these techniques with hysteroscopy, and to determine which SCSH findings best discriminate between benign and malignant endometrium.

METHODS

From January 2003 to December 2005, consecutive women referred to our ultrasound unit because of post-menopausal bleeding and who were found to have endometrial thickness ≥ 4.5 mm as measured by trans-vaginal sonography underwent SCSH and were enrolled into the study. The Ethics Committee of Lund University approved the study protocol. Informed consent was obtained from all participants after the nature of the procedures had been fully explained.

The ultrasound examinations were carried out by one of two examiners (L.V., P.S.) using a GE Voluson 730 Expert ultrasound system (GE Medical Systems, Zipf, Austria) with a 2.8–10-MHz transvaginal transducer. Both examiners had more than 15 years' experience with 2D gynecological ultrasound imaging, but one of them was less experienced with 3D gynecological ultrasound examination than the other (1 year vs. 5 years). Endometrial thickness was measured and SCSH was carried out as described previously[2,8]; during SCSH the whole uterine cavity was meticulously and systematically scanned from one side to the other on a sagittal section through the uterus, and from the bottom to the top of the uterine cavity on a transverse section through the uterus. If there were problems with back flow, more fluid was infused (in some cases almost continuously) to keep the cavity expanded during the examination. The 2D SCSH ultrasound images were videotaped (duration of video recording approximately 1 min). After videotaping, a 3D volume of the uterine corpus was acquired; a longitudinal view of a satisfactory gray-scale image of the uterus was obtained and the uterine corpus was centered within the 3D sector on the screen so that it filled the whole sector. Using the setting 'Quality high 2' for 3D sweep speed, a volume of the uterine corpus with fluid in the cavity was acquired. The resultant multiplanar display was examined to ensure that a complete volume of the uterine corpus with its fluid-filled cavity had been captured. In cases where more than one volume of the uterine corpus had been acquired only the best one was stored electronically for later analysis.

After SCSH an endometrial sample was taken using the Endorette® endometrial sampling device (Medscand AB, Malmö, Sweden). The Endorette is a sterile curette with a polyethylene piston which slides within a straight but flexible polypropylene sheath with four lateral holes near its tip. Its length is 285 mm and its outer diameter

is 2.6 mm. Unless the Endorette sample revealed a malignancy, women with focal lesions detected at SCSH were recommended to undergo hysteroscopic resection of the focal lesion(s) under general anesthesia. Hysteroscopy was performed by any of six gynecologists specifically trained to perform this procedure. Women with no focal lesions were recommended to have dilatation and curettage. A final diagnosis was determined on the basis of histological examination of surgical specimens obtained by dilatation and curettage, hysteroscopic resection or hysterectomy. Staging of malignant tumors was done by the attending physician in accordance with the classification system recommended by the International Federation of Gynecology and Obstetrics[9].

After completion of the scanning phase of the study, the videotapes and 3D ultrasound volumes were analyzed independently by two examiners (L.V., P.S.). The 3D ultrasound volumes were analyzed using the Z technique[10] and the orientation guidelines described by Merz *et al.*[11]. A standardized research protocol was followed when evaluating the ultrasound findings. For women who underwent hysteroscopy, a similar standardized form was filled in by the hysteroscopist immediately after the hysteroscopy. The hysteroscopist had no knowledge of the ultrasound findings when performing the hysteroscopy. The following features were evaluated by the ultrasound examiner and the hysteroscopist: presence, number and surface regularity of focal lesions or, in the absence of focal lesions, surface regularity of the endometrium (the latter was not evaluated by the hysteroscopist). The definition of a focal lesion was 'any visible focal irregularity protruding into the uterine cavity, either a localized protrusion of the endometrium or a localized protrusion of other tissue, e.g. a submucosal myoma'. The definition of a uterine cavity without a focal lesion was 'a cavity with completely smooth internal contour'. When judging surface regularity, an undulating surface was regarded as smooth, whereas a 'spiky' surface was regarded as irregular. The location of focal lesions was also described (anterior, posterior, right, left, upper, middle, lower or fundal part of the uterine cavity). Agreement with regard to location of the focal lesions was classified as agreed (i.e. agreement for all focal lesions for which agreement was assessable; agreement could not be assessed for all lesions if there was an interobserver difference in number of focal lesions, for example if one observer stated that there were two focal lesions in the uterine cavity and the other stated that there were three, agreement with regard to location was assessable for two lesions), partially agreed (i.e. agreement for some but not all focal lesions for which agreement was assessable), disagreed (i.e. disagreement for all focal lesions for which agreement was assessable) or not possible to assess (agreement could not be assessed if at least one observer suggested that there were no focal lesions or found that localization was not possible, for example because of myomas distorting or shadowing the uterine cavity). After independently assessing all videotapes and 3D ultrasound volumes, the two ultrasound observers

reviewed their completed paper forms together. If there was disagreement with regard to any of the variables they went back to the videotapes/volumes and reviewed them together to reach a consensus. Their agreed description (consensus opinion) of the uterine cavity was compared with the final histopathological diagnosis and with the hysteroscopy findings.

Exclusion criteria were: failed SCSH, incomplete videotaping, absence of histopathological diagnosis or insufficient endometrial material for diagnosis, histopathological diagnosis obtained only by Endorette, or both examiners finding it impossible to reliably evaluate the uterine cavity. Patients with a final diagnosis based on Endorette samples were excluded because we wanted to ensure that only representative samples were used to establish the final diagnosis[12].

A woman was considered to be postmenopausal if she reported absence of menstruation for at least 1 year after the age of 40 years provided that the amenorrhea was not explained by pregnancy, medication or disease. Postmenopausal bleeding was defined as any vaginal bleeding in a postmenopausal woman not on hormone replacement therapy, or unscheduled vaginal bleeding in a postmenopausal woman receiving hormonal replacement therapy.

Statistical analysis

Statistical calculations were undertaken using the Statistical Package for the Social Sciences version 12.02 (SPSS Inc., Chicago, IL, USA). The statistical significance of a difference in unpaired continuous data was calculated using the Mann–Whitney U-test (because the data were not normally distributed) and that of a difference in unpaired discrete data using Fisher's exact test. Two-tailed $P < 0.05$ was considered statistically significant.

Agreement analyses

Percentage agreement between observers and between SCSH and hysteroscopy was determined and, when possible, a kappa value was calculated. Kappa values are a measure of how much the observed agreement exceeds agreement by chance and can be calculated only if field tables are symmetrical. Kappa values will be low if data are very skewed, even if agreement is close to 100% (e.g. if two observers agree that a particular finding is present in 90% of cases and absent in 5% of cases)[13].

The difference in number of focal lesions between observers and the difference in number of focal lesions between SCSH and hysteroscopy were also calculated (distribution of absolute differences, mean difference ± 2 SD, and the 95% CI of the mean difference; if the 95% CI of the mean difference included zero it was assumed that no systematic bias between observers or methods existed).

The McNemar test was used to determine whether differences in agreement between 2D and 3D SCSH

were statistically significant, two-tailed $P < 0.05$ being considered statistically significant.

Diagnostic performance with regard to discrimination between benign and malignant endometrium

The statistical significance of a possible relationship between single ultrasound variables and endometrial malignancy was determined using univariate logistic regression with the likelihood ratio test, with two-tailed $P < 0.05$ being considered statistically significant.

Receiver–operating characteristics (ROC) curves were drawn for each ultrasound variable, and the area under the ROC curve (AUC) and the 95% CI of this area were calculated. If the lower limit of the CI for the AUC was > 0.5 then the variable was considered to have discriminatory potential. For number of focal lesions the ROC curve was also used to determine mathematically the best cut-off value for number of focal lesions to predict malignancy, with this value being defined as that corresponding to the point on the ROC curve situated farthest from the reference line. The sensitivity, specificity, positive and negative likelihood ratios (LR+ and LR−, respectively) of the ultrasound variables with regard to endometrial malignancy were calculated. We defined the best diagnostic test as the one with the largest AUC, the highest LR+ and the lowest LR−. The McNemar test was used to determine whether a difference in sensitivity and specificity between 2D and 3D SCSH results was statistically significant, two-tailed $P < 0.05$ being considered statistically significant.

Multivariate logistic regression was used to test whether combining endometrial thickness with the best SCSH variable would improve discrimination between benign and malignant endometrium, with the likelihood ratio test yielding $P < 0.05$ being the criterion for including the additional variable in the model.

RESULTS

One hundred and seventy consecutive women referred to our ultrasound unit because of postmenopausal bleeding, and found to have endometrial thickness ≥ 4.5 mm at transvaginal ultrasound examination, underwent 2D and 3D SCSH and were enrolled in the study. Of these, 86 women were excluded for the following reasons: no histological diagnosis ($n = 16$) or diagnosis obtained by Endorette only ($n = 12$), either 2D or 3D SCSH ultrasound images impossible to assess owing to technical problems, e.g. shadowing from myomas or insufficient intracavitary fluid ($n = 10$; in five cases neither the 2D nor the 3D SCSH images could be evaluated, in another five cases only the 3D SCSH could not be evaluated), failed SCSH, e.g. because of cervical stenosis or complete back flow ($n = 29$), and SCSH not properly videotaped ($n = 19$). The characteristics of the 84 women included and the 86 women excluded are shown in Table 1.

Of the 84 women included, 70 had benign and 14 (17%) had malignant endometrium. The age of

Table 1 Characteristics of patients included in and excluded from further analysis in the study

Variable	Included (n = 84)	Excluded (n = 86)
Age (years)	66 (50–89)	67 (50–93)
Hormone replacement therapy	29 (35)	35 (41)
Endometrial thickness (mm)	11.3 (4.6–41.0)	9.6 (4.5–39.7)
Final diagnosis		
Benign	70 (83)	42 (49)
Malignant	14 (17)	16 (19)*
Benign diagnosis by Endorette only	—	12 (14)
Malignant diagnosis by Endorette only	—	0 (0)
Material insufficient for diagnosis by Endorette only	—	7 (8)
Material insufficient for diagnosis by D&C only	—	2 (2)
No diagnostic procedure performed	—	7 (8)
Method of final diagnosis		
Hysteroscopy	55 (65)†	32 (37)
Hysterectomy	20 (24)	20 (23)
D&C	9 (11)	8 (9)
Endorette sampling	—	19 (22)
No diagnostic procedure performed	—	7 (8)
Histological diagnosis		
Benign		
Polyp	42 (50)	25 (29)
Myoma	7 (8)	2 (2)
Atrophic endometrium	12 (14)	11 (13)
Estrogen-influenced endometrium	3 (4)	5 (6)
Gestagen-influenced endometrium	0 (0)	1 (1)
Hyperplasia		
Without atypia	4 (5)	8 (9)
With atypia	2 (2)	1 (1)
Pyometra	0 (0)	1 (1)
Malignant		
Adenocarcinoma	12 (14)	15 (17)
Carcinosarcoma	2 (2)	1 (1)
No histological diagnosis	—	16 (19)
Stage of malignant tumors‡		
I	10 (71)	11 (69)
II	3 (21)	1 (6)
III	1 (7)	3 (19)
Not staged	0 (0)	1 (6)

Data are expressed as median (range) or *n* (%). *Reasons for exclusion were failed saline contrast sonohysterography (SCSH) (*n* = 9), inability to evaluate the uterine cavity both at two-dimensional (2D) and three-dimensional SCSH (*n* = 5), and 2D SCSH not videotaped properly (*n* = 2). †Fifty-six patients underwent hysteroscopy, but one was diagnosed finally by hysterectomy. ‡Percentages calculated as proportion of malignant tumors in that group. D&C, dilatation and curettage.

the women with malignant endometrium was similar to that of those with benign endometrium (median, 68 (range, 56–82) years vs. 64 (range, 50–89) years; $P = 0.334$, Mann–Whitney U-test), but women with malignant endometrium had a thicker endometrium (median, 16.1 (range, 8.2–41.0) mm vs. 10.3 (range, 4.6–36.0) mm; $P = 0.001$, Mann–Whitney U-test), and fewer used hormone replacement therapy (1/14 (7%) vs. 28/70 (40%); $P = 0.028$, Fisher's exact test).

Interobserver agreement for saline contrast sonohysterography

Interobserver agreement with regard to presence of focal lesions and presence of irregular focal lesions was similar (88–98%) for 2D and 3D SCSH. Interobserver agreement for number and location of focal lesions was also similar for 2D and 3D SCSH (63–76% for number and 70–77%

for location), but it was poorer than that for presence of focal lesions. Details are shown in Table 2.

Agreement between saline contrast sonohysterography and hysteroscopy

In two of the total of 56 patients who underwent hysteroscopy, the hysteroscopist did not fill in the standard form. The information from the remaining 54 women was used to determine the agreement between SCSH and hysteroscopy. One of the 54 women had endometrial cancer. In this woman, the ultrasound observers noted three irregular focal lesions on 2D SCSH but no focal lesions on 3D SCSH. At hysteroscopy six irregular focal lesions were seen. The Endorette biopsy showed sparse endometrium without malignant changes, but hysteroscopic resection showed endometrial hyperplasia with complex atypia. The

Table 2 Interobserver agreement for saline contrast sonohysterography (SCSH) results

	2D SCSH		3D SCSH	
Feature	% agreement (n)	Kappa	% agreement (n)	Kappa
Results for all 84 patients				
Interobserver agreement				
Presence of focal lesions (yes/no)	95 (80/84)	0.77	89 (75/84)	0.68
Irregular focal lesion (yes/no)	89 (75/84)	0.55	88 (74/84)	0.38
Irregular uterine cavity (yes/no)	88 (74/84)	0.51	88 (74/84)	0.38
Number of focal lesions (0, 1, 2, ≥ 3)	67 (54/81*)	0.52	63 (51/81*)	0.49
Location of focal lesion†				
Agreement	77 (46/60‡)	N/A	70 (42/60‡)	N/A
Partial agreement	12 (7/60‡)	N/A	17 (10/60‡)	N/A
Disagreement	12 (7/60‡)	N/A	13 (8/60‡)	N/A
Not possible to assess	15 (13/84)§	N/A	27 (23/84)§	N/A
Results for 54 patients who underwent hysteroscopy				
Interobserver agreement				
Presence of focal lesions (yes/no)	98 (53/54)	—¶	94 (51/54)	0.64
Irregular focal lesion (yes/no)	93 (50/54)	0.30**	98 (53/54)	0.66
Irregular uterine cavity (yes/no)	93 (50/54)	0.30**	98 (53/54)	0.66
Number of focal lesions (0, 1, 2, ≥ 3)	76 (41/54)	—¶	63 (34/54)	0.45
Location of focal lesion†				
Agreement	72 (33/46‡)	N/A	70 (32/46‡)	N/A
Partial agreement	15 (7/46‡)	N/A	15 (7/46‡)	N/A
Disagreement	13 (6/46‡)	N/A	15 (7/46‡)	N/A
Not possible to assess	4 (2/54)	N/A	13 (7/54)	N/A

*In three cases it was not possible to determine the number of focal lesions. †Definitions of 'agreement', 'partial agreement', 'disagreement' and 'not possible to assess' are given in the methods section. ‡Number of cases in which the location of focal lesions was assessable for both two-dimensional (2D) and three-dimensional (3D) SCSH. §The only difference in agreement that was statistically significant ($P = 0.006$). ¶Absent kappa values are explained by asymmetric field tables making it impossible to calculate kappa values. **The low kappa value (0.30) associated with a high percentage agreement (93%) is explained by highly skewed data. N/A, not applicable.

Table 3 Saline contrast sonohysterography (SCSH) and hysteroscopy findings in the 54 patients used to determine agreement between SCSH findings and hysteroscopy findings

	2D SCSH			3D SCSH			
Feature	Obs. 1	Obs. 2	Consensus opinion	Obs. 1	Obs. 2	Consensus opinion	Hysteroscopy
Presence of focal lesions	54	53	54	49	50	49	51
Presence of irregular focal lesions	2	4	2	2	1	1	12
Irregular cavity	2	4	2	2	1	1	N/A
Number of focal lesions							
0	0	1	0	5	4	5	3
1	26	26	28	27	22	23	23
2	19	18	16	16	16	16	19
≥ 3	9	9	10	6	12	10	9

Data are expressed as *n*. 2D, two dimensional; 3D, three dimensional; N/A, not applicable; Obs., Observer.

hysterectomy specimen revealed a highly differentiated Stage I endometrial adenocarcinoma.

SCSH and hysteroscopy findings in the 54 women who underwent hysteroscopy are shown in Table 3, and agreement of 2D and 3D SCSH with hysteroscopy in Table 4. The agreement between 2D SCSH and hysteroscopy with regard to presence of focal lesions, presence of focal lesions with irregular surface and location of focal lesions was similar to that between 3D SCSH and hysteroscopy (94% vs. 93%, 74% vs. 76% and 66% vs. 64%, respectively), whereas agreement with regard to number of focal lesions tended to be

superior for 2D SCSH (63% vs. 54%) but not statistically significantly so ($P = 0.39$, McNemar test). One observer recorded fewer focal lesions at 3D SCSH than the other. No other systematic differences with regard to number of focal lesions between observers or between methods were found (Table S1).

Diagnostic performance of saline contrast sonohysterography

The sensitivity, specificity, LR+ and LR−, and AUC for various ultrasound variables with regard to endometrial

Table 4 Agreement between saline contrast sonohysterography (SCSH) and hysteroscopy results

	2D SCSH		3D SCSH	
Feature	*% agreement*	*Kappa*	*% agreement*	*Kappa*
Presence of focal lesions (yes/no)	94 (51/54)	—*	93 (50/54)	0.46
Irregular focal lesion (yes/no)	74 (40/54)	−0.07†	76 (41/54)	−0.04†
Number of focal lesions (0, 1, 2, ≥ 3)	63 (34/54)	—*	54 (29/54)	0.32
Location of focal lesion‡				
Agreement	66 (29/44§)	N/A	64 (28/44§)	N/A
Partial agreement	16 (7/44§)	N/A	14 (6/44§)	N/A
Disagreement	18 (8/44§)	N/A	23 (10/44§)	N/A
Not possible to assess	13 (7/54)	N/A	17 (9/54)	N/A

The consensus opinion of two observers was compared with hysteroscopy results. *Absent kappa values are explained by asymmetric field tables making it impossible to calculate kappa values. †Low kappa values (−0.07 and −0.04) associated with a relatively high percentage agreement (74% and 76%) are explained by highly skewed data. ‡Definitions of 'agreement', 'partial agreement', 'disagreement' and 'not possible to assess' are given in the methods section. §Number of cases in which the location of focal lesions was assessable for both two-dimensional (2D) and three-dimensional (3D) SCSH. None of the differences in agreement was statistically significant. N/A, not applicable.

Table 5 Diagnostic performance of ultrasound variables with regard to endometrial malignancy

Feature	*AUC (95% CI)*	*Optimal cut-off*	*Sensitivity (% (n))*	*Specificity (% (n))*	*LR+*	*LR−*	*P**
Endometrial thickness	0.78 (0.68–0.89)	12.5 mm	93 (13/14)	69 (48/70)	3.0	0.1	0.013
2D SCSH†							
Focal lesions (yes/no)	0.57 (0.42–0.72)	—	100 (14/14)	14 (10/70)	1.2	—	0.048
Irregular focal lesion (yes/no)	0.84 (0.70–0.99)	—	71 (10/14)	97 (68/70)	25.0	0.3	0.0005
Irregular uterine cavity‡ (yes/no)	0.83 (0.68–0.97)	—	71 (10/14)	94 (66/70)	12.5	0.3	0.0005
Number of focal lesions§	0.65 (0.50–0.80)	≥ 2	—	—	—	—	—
Number of focal lesions (≥ 2¶ vs. 0−1)	0.61 (0.45–0.78)		64 (9/14)	59 (41/70)	1.6	0.6	0.117
3D SCSH†							
Focal lesions (yes/no)	0.53 (0.37–0.69)	—	86 (12/14)	20 (14/70)	1.1	0.7	0.609
Irregular focal lesion (yes/no)	0.70 (0.52–0.88)	—	43 (6/14)	97 (68/70)	15.0	0.6	0.0005
Irregular uterine cavity‡ (yes/no)	0.69 (0.52–0.87)	—	43 (6/14)	96 (67/70)	9.9	0.6	0.0005
Number of focal lesions§	0.61 (0.44–0.79)	≥ 3	—	—	—	—	—
Number of focal lesions (≥ 3¶ vs. 0−2)	0.59 (0.42–0.77)		36 (5/14)	83 (58/70)	2.1	0.8	0.135

*Univariate logistic regression with likelihood ratio test. †Consensus opinion of two observers. ‡Presence of irregular focal lesion or, in the absence of a focal lesion, irregular surface of the endometrium. §Receiver−operating characteristics (ROC) curve constructed using all values for number of focal lesions. ¶Optimal cut-off for number of focal lesions. 2D, two dimensional; 3D, three dimensional; LR, likelihood ratio; AUC, area under ROC curve; SCSH, saline contrast sonohysterography.

malignancy are shown in Table 5. The number of focal lesions did not discriminate between benign and malignant endometrium. For both 2D and 3D SCSH the ultrasound variable that best discriminated between benign and malignant endometrium was the presence of at least one focal lesion with irregular surface (Figure 1): AUC 0.84 (95% CI, 0.70–0.99), sensitivity 71% (95% CI, 42–92%), specificity 97% (95% CI, 90–100%), LR+ 25.0 (95% CI, 4.6–1044) and LR− 0.3 (95% CI, 0.05–0.61) for 2D SCSH; and AUC 0.70 (95% CI, 0.52–0.88), sensitivity 43% (95% CI, 18–71%), specificity 97% (95% CI, 90–100%), LR+ 15 (95% CI, 1.7–814) and LR− 0.6 (95% CI 0.22–0.89) for 3D SCSH. The large difference in sensitivity between 2D and 3D SCSH was not statistically significant ($P = 0.13$, McNemar test).

Multivariate logistic regression showed that adding endometrial thickness to the best SCSH variable (irregular focal lesion) did not improve discrimination between benign and malignant endometrium.

DISCUSSION

Both interobserver agreement and agreement with hysteroscopy was similar for 2D and 3D SCSH with high percentage agreement for presence of focal lesions and presence of irregular focal lesions, but poorer agreement with regard to number and location of focal lesions. We cannot claim to have shown 2D and 3D SCSH to be equivalent diagnostic methods, but there was nothing to indicate that 3D SCSH was superior to 2D SCSH either in terms of interobserver reproducibility, agreement with hysteroscopy findings or correctly diagnosing endometrial malignancy. There were more 3D examinations than 2D examinations for which it was not possible to determine interobserver agreement with regard to location
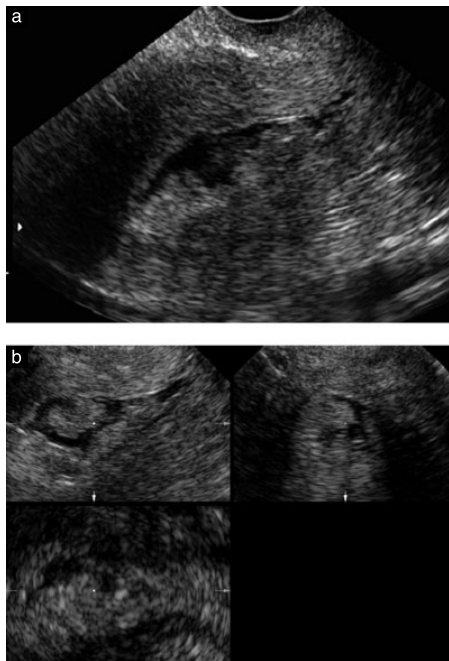
Figure 1 Two-dimensional (2D) (a) and three-dimensional (b) saline contrast sonohysterography (SCSH) ultrasound images illustrating irregular surface of focal lesions. Both (a) and (b) are from the same patient. The histopathological diagnosis was adenocarcinoma. The images illustrate the difficulties that one may encounter when trying to determine the number of focal lesions. Image (a) can be interpreted as showing either one focal lesion or two focal lesions with one fusing into the other. Clearly, the number of focal lesions cannot be determined from one single image. In this particular case the observers agreed that there were three focal lesions on 2D SCSH when the whole cavity was scanned meticulously and systematically from one side to the other and from the bottom to the top of the uterine cavity, and on there being four focal lesions seen in the acquired volume analyzed using the Z technique.

of focal lesions, and there were more 3D examinations than 2D examinations in which the uterine cavity could not be evaluated at all because of technical problems, e.g. myomas casting shadows or distorting the image (10 vs. 5). The latter might be explained by a videotape of a live scan offering greater possibilities for evaluating the cavity, because during a live scan the ultrasound probe is moved continuously to improve visualization of different parts of the uterine cavity, whereas a 3D ultrasound volume consists of a series of still images that have been acquired automatically, and some images in the volume may not have been acquired under optimal conditions. Another explanation could be that, because the video recording of the 2D SCSH was done before the 3D volume was acquired, there is a theoretical possibility

that there could have been less fluid in the uterine cavity when the 3D volumes were acquired than when the 2D scans were videotaped, even though we infused more fluid when we had problems with back flow. However, it rarely happened that back flow was a problem specifically during the acquisition of the volume, and we do not believe that the order of the examinations affected our results. It is possible that back-flow problems could be overcome if gel were used instead of saline[14].

One difficulty when performing an agreement study is to provide exact definitions of the outcome variables. For example, judging the regularity of the surface of a focal lesion or of the endometrium is completely subjective, even if definitions are provided. Defining what constitutes a focal lesion is also problematic. When there are protrusions/focal lesions close to each other, there may be difficulties in determining the number of focal lesions, as the findings could be interpreted as only one focal lesion, as many focal lesions with one fusing into the other, as a very irregular endometrium (Figure 1) or as endometrial folds. To the best of our knowledge there is no generally accepted definition of what constitutes a focal lesion in the uterine cavity, nor do we know of a generally accepted definition of 'irregular surface'.

A limitation of our study is that half of the women in this consecutive series of patients needed to be excluded. However, the patients included and excluded had similar demographic data, a similar rate of endometrial cancer and similar endometrial cancer stages. Therefore, we believe that the patients included are indeed representative of the total series of consecutive patients with postmenopausal bleeding and endometrial thickness ≥ 4.5 mm who were referred to us. Most of the exclusions were because there was no reliable final histological diagnosis or because of technical problems at SCSH (suboptimal filling of the uterine cavity or obscured view of the uterine cavity because of myomas, etc.). In a meta-analysis the failure rate of SCSH in postmenopausal women was stated to be 13.5% (95% CI, 10.2–16.8%)[15]. Our failure rate of 17% (29/170) was similar to the upper value of the 95% CI of this rate. A second limitation of our study is that the 54 women who underwent hysteroscopy constitute a rather homogeneous study sample with regard to presence of focal lesions and irregular focal lesions (Table 3). This means that four-field tables illustrating agreement are likely to be asymmetrical, and in this situation a kappa value will be a poor indicator of agreement[13]. A third limitation is that hysteroscopy was performed after an Endorette sample had been taken. This could have affected the agreement between SCSH and hysteroscopy findings. We cannot exclude the possibility that the endometrial biopsy removed some focal lesions at least partially, but it is highly unlikely that any focal lesion was removed totally[12,16]. It is more likely that the surface of a focal lesion could have been changed by the biopsy, and therefore that the biopsy could have affected the agreement between SCSH and hysteroscopy with regard to surface regularity.

To the best of our knowledge, there are no published studies comparing interobserver agreement of 2D SCSH with that of 3D SCSH. In the only three publications that compared 2D and 3D SCSH with regard to their agreement with hysteroscopy, the agreement with hysteroscopy findings was similar for 2D and 3D SCSH[4,5,7]. However, the results are not directly comparable with ours because of differences in study populations, study design and outcome variables.

Irregular surface of the endometrium was mentioned to be a sign of endometrial malignancy in a few publications[3,17–19], but the diagnostic performance of irregular focal lesions was reported in only one study (sensitivity 88%, specificity 66%, LR+ 2.58 and LR− 0.18)[3]. In our study the presence of irregular focal lesions increased the odds of malignancy substantially (25 times and 15 times for 2D and 3D SCSH, respectively), whereas the absence of irregular focal lesions decreased the odds only a little, and the difference in diagnostic performance between 2D and 3D SCSH was not statistically significant.

It is often emphasized that one advantage of 3D over 2D ultrasound imaging is that a volume can be stored and analyzed later as many times as needed. However, 2D examinations can be stored as electronic clips or on videotape, and clips and videotapes can also be analyzed repeatedly. A quick consultation for a second opinion by sending volumes over the internet is not yet possible, because 3D volumes are currently too 'heavy' to be sent over the internet, whereas still images and video clips can be sent. 3D ultrasound examination offers the possibility of studying an infinite number of sections through the uterine cavity. This is a potential advantage, but in our study this advantage did not improve either the interobserver reproducibility of ultrasound findings, agreement with hysteroscopy findings, or sensitivity and specificity with regard to diagnosing endometrial cancer. 3D SCSH is at least as operator dependent as 2D SCSH; to obtain a good 3D volume the 2D image needs to be optimized, the acquisition of the volume must be optimal, e.g. the uterine corpus must be placed in the center of the 3D 'box', and an optimal angle of the volume must be used. Analysis of the 3D volume is also highly operator dependent. For example, orientation is definitely not intuitive but requires experience and practice[11]. Had both observers been extremely experienced in manipulating the 3D volumes, our results might have been different. We also believe that 2D and 3D SCSH are equally time consuming for the ultrasound examiner. Even though the acquisition of a volume does not take long, it takes a while to optimize and acquire the image, and analyzing 3D volumes is at least as time consuming as evaluating ultrasound findings during a live scan.

In conclusion, our results do not suggest that 3D SCSH is superior to 2D SCSH performed by experienced examiners either in terms of interobserver agreement, agreement with hysteroscopy findings or for correctly diagnosing endometrial malignancy. The presence of an irregular focal lesion in the uterine cavity seems to be the best SCSH variable for discrimination between benign and malignant endometrium in women with postmenopausal bleeding and endometrial thickness ≥ 4.5 mm.

REFERENCES

1. Smith-Bindman R, Kerlikowske K, Feldstein VA, Subak L, Scheidler J, Segal M, Brand R, Grady D. Endovaginal ultrasound to exclude endometrial cancer and other endometrial abnormalities. *JAMA* 1998; **280**: 1510–1517.
2. Epstein E, Ramirez A, Skoog L, Valentin L. Transvaginal sonography, saline contrast sonohysterography and hysteroscopy for the investigation of women with postmenopausal bleeding and endometrium > 5 mm. *Ultrasound Obstet Gynecol* 2001; **18**: 157–162.
3. Epstein E, Valentin L. Gray-scale ultrasound morphology in the presence or absence of intrauterine fluid and vascularity as assessed by color Doppler for discrimination between benign and malignant endometrium in women with postmenopausal bleeding. *Ultrasound Obstet Gynecol* 2006; **28**: 89–95.
4. Bonilla-Musoles F, Raga F, Osborne NG, Blanes J, Coelho F. Three-dimensional hysterosonography for the study of endometrial tumors: comparison with conventional transvaginal sonography, hysterosalpingography, and hysteroscopy. *Gynecol Oncol* 1997; **65**: 245–252.
5. La Torre R, De Felice C, De Angelis C, Coacci F, Mastrone M, Cosmi EV. Transvaginal sonographic evaluation of endometrial polyps: a comparison with two dimensional and three dimensional contrast sonography. *Clin Exp Obstet Gynecol* 1999; **26**: 171–173.
6. Dueholm M, Lundorf E, Sorensen JS, Ledertoug S, Olesen F, Laursen H. Reproducibility of evaluation of the uterus by transvaginal sonography, hysterosonographic examination, hysteroscopy and magnetic resonance imaging. *Hum Reprod* 2002; **17**: 195–200.
7. de Kroon CD, Louwe LA, Trimbos JB, Jansen FW. The clinical value of 3-dimensional saline infusion sonography in addition to 2-dimensional saline infusion sonography in women with abnormal uterine bleeding: work in progress. *J Ultrasound Med* 2004; **23**: 1433–1440.
8. Sladkevicius P, Valentin L, Marsal K. Endometrial thickness and Doppler velocimetry of the uterine arteries as discriminators of endometrial status in women with postmenopausal bleeding: a comparative study. *Am J Obstet Gynecol* 1994; **171**: 722–728.
9. Shepherd JH. Revised FIGO staging for gynaecological cancer. *Br J Obstet Gynaecol* 1989; **96**: 889–892.
10. Abuhamad AZ, Singleton S, Zhao Y, Bocca S. The Z technique: an easy approach to the display of the mid-coronal plane of the uterus in volume sonography. *J Ultrasound Med* 2006; **25**: 607–612.
11. Merz E, Benoit B, Blaas HG, Baba K, Kratochwil A, Nelson T, Pretorius D, Jurkovic D, Chang FM, Lee A; ISUOG 3D Focus

Group. Standardization of three-dimensional images in obstetrics and gynecology: consensus statement. *Ultrasound Obstet Gynecol* 2007; **29**: 697–703.

12. Epstein E, Skoog L, Valentin L. Comparison of Endorette and dilatation and curettage for sampling of the endometrium in women with postmenopausal bleeding. *Acta Obstet Gynecol Scand* 2001; **80**: 959–964.

13. Kundel HL, Polansky M. Measurement of observer agreement. *Radiology* 2003; **228**: 303–308.

14. Exalto N, Stappers C, van Raamsdonk LA, Emanuel MH. Gel instillation sonohysterography: first experience with a new technique. *Fertil Steril* 2007; **87**: 152–155.

15. de Kroon CD, de Bock GH, Dieben SW, Jansen FW. Saline contrast hysterosonography in abnormal uterine bleeding: a systematic review and meta-analysis. *BJOG* 2003; **110**: 938–947.

16. Epstein E, Ramirez A, Skoog L, Valentin L. Dilatation and curettage fails to detect most focal lesions in the uterine cavity in women with postmenopausal bleeding. *Acta Obstet Gynecol Scand* 2001; **80**: 1131–1136.

17. Bronz L, Suter T, Rusca T. The value of transvaginal sonography with and without saline instillation in the diagnosis of uterine pathology in pre- and postmenopausal women with abnormal bleeding or suspect sonographic findings. *Ultrasound Obstet Gynecol* 1997; **9**: 53–58.

18. Laifer-Narin SL, Ragavendra N, Lu DS, Sayre J, Perrella RR, Grant EG. Transvaginal saline hysterosonography: characteristics distinguishing malignant and various benign conditions. *AJR Am J Roentgenol* 1999; **172**: 1513–1520.

19. Davis PC, O'Neill MJ, Yoder IC, Lee SI, Mueller PR. Sonohysterographic findings of endometrial and subendometrial conditions. *Radiographics* 2002; **22**: 803–816.

SUPPORTING INFORMATION ON THE INTERNET

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

Table S1 Distribution of interobserver differences and differences between saline contrast sonohysterography (SCSH) and hysteroscopy with regard to number of focal lesions.

Table S1 Distribution of interobserver differences and differences between saline contrast sonohysterography (SCSH) and hysteroscopy with regard to number of focal lesions.

	Difference		Distribution of absolute differences			
	Mean	95% CI Limits of agreement	0 n (%)	1 n (%)	2 n (%)	≥3 n (%)
Interobserver difference for 81[*] patients						
2D hydrosonography	-0.01	-0.18 – 0.16	49 (61)	27 (33)	5 (6)	0
3D hydrosonography	-0.38	-0.60 – -0.16	48 (59)	23 (29)	8 (10)	2 (2)
Interobserver difference for 54 patients who underwent hysteroscopy						
2D hydrosonography	0.04	-0.11 – 0.19	41 (76)	12 (22)	1 (2)	0
3D hydrosonography	-0.24	-0.44 – -0.04	34 (63)	18 (33)	1 (2)	1 (2)
Difference between hydrosonography and hysteroscopy[†] (n = 54)						
2D hydrosonography vs. hysteroscopy	-0.24	-0.67 – 0.19	32 (59)	13 (24)	4 (8)	5 (9)
3D hydrosonography vs. hysteroscopy	-0.33	-0.78 – 0.11	28 (52)	19 (35)	2 (4)	5 (9)

2D, two-dimensional; 3D, three-dimensional
[*] 81 patients (not 84) because in three cases one of the observers was unable to state the number of focal lesions
[†] The number seen at hydrosonography (consensus opinion of two observers) minus the number seen at hysteroscopy