Building decision trees for diagnosing intracavitary uterine pathology

T. Van den Bosch[1], A. Daemen[2], O. Gevaert[2], B. De Moor[2], D. Timmerman[1]

[1]Department of Obstetrics and Gynaecology, University Hospitals K.U.Leuven, 3000 Leuven, Belgium.
[2]Department of Electrical Engineering, ESAT-SCD, K.U.Leuven, 3000 Leuven, Belgium.

Correspondence at: thierry.van.den.bosch@skynet.be

Abstract

Objectives: To build decision trees to predict intrauterine disease, based on a clinical data set, and using mathematical software.

Methods: Diagnostic algorithms were built and validated using the data of 402 consecutive patients who underwent grey scale ultrasound, followed by colour Doppler, saline infusion sonography (SIS), office hysteroscopy and endometrial sampling. The "final diagnosis" was classified as "abnormal" in case of endometrial polyps, hyperplasia or malignancy or intracavitary myoma. "Pre-test parameters" included patient's age, weight, length, parity, menopausal status, bleeding symptoms and cervical cytology; "post-test parameters" included ultrasound-, color Doppler-, SIS-, hysteroscopy- findings and histology results after endometrial sampling. Decision Tree #1 was built using both "pre-test" and "post-test" parameters; Tree #2 was only based on "post-test" parameters; Tree #3 was designed without using the hysteroscopy variables. The Waikato Environment for Knowledge Analysis (Weka) software was used for the development of decision trees.

Results: All trees started with an imaging technique: hysteroscopy or SIS. The diagnostic accuracy was 88.3%, 88.3% and 84.0% for Tree #1, #2 and #3 respectively, the sensitivity and specificity was 95.5% and 82%, 97.7% and 80.0, 93.2 and 76.0%, respectively.

Conclusion: The method used in this study enables the comparison between different decision trees containing multiple tests.

Key words: Decision trees, SIS, endometrial sampling, ultrasound, hysteroscopy, abnormal uterine bleeding.

Introduction

The question which diagnostic test should be used in case of abnormal uterine bleeding is still a matter of debate. Diagnostic algorithms for intrauterine disease are usually based on studies and meta-analyses published in the literature and mostly include one or more of the following diagnostic modalities: office endometrial sampling (Dijkhuizen *et al*., 2000; Clark *et al*., 2002a), hysteroscopy (Clark *et al*., 2002b), ultrasound (Smith-Bindman *et al*., 1998; Gupta *et al*., 2002; Tabor *et al*., 2002), fluid contrast sono-hysterography (SHG) (de Kroon *et al*., 2003). Office endometrial sampling is accurate in the diagnosis of endometrial cancer (Clark *et al*., 2002a), but misses most focal lesions, such as polyps (Van den Bosch *et al*., 1995). Hysteroscopy is considered the gold standard to diagnose focal intracavity lesions, but

performs somewhat less in the detection of malignancy (Clark *et al*., 2002b). Ultrasound is useful in the triage of postmenopausal patients at risk for endometrial disease, by measuring the endometrial thickness (Tabor *et al*., 2002; Smith-Bindman *et al*., 1998). Sonohysterography (SHG) has been proposed as first step examination in the diagnosis of focal lesions such as endometrial polyps and intracavity fibroids (de Kroon *et al*., 2003). The choice of a diagnostic examination may also be influenced by the personal skills and preference of the clinician, as well as by the availability of the diagnostic tools.

A decision tree is a supervised approach to classification. It is a tree data structure constructed from a set of patients. Each patient is described by a set of attributes: variables with numeric (e.g. endometrium thickness in mm) or symbolic values (e.g. "menopausal" or "premenopausal"). Each non-

terminal node of a decision tree contains a test on one or more attributes (e.g. SIS) which result (e.g. "lesion" or "no lesion") is used to select the branch to follow from that node. The terminal nodes reflect the decision outcomes (e.g. "normal" or "abnormal") (Quinlan, 1993).

In this study, we built decision trees to predict intrauterine disease, based on a clinical data set, and using mathematical software.

Methods

A dataset of 402 consecutive patients evaluated between October 2004 and November 2006 at the One Stop Bleeding Clinic of the University Hospital Leuven were included. The data used have been extensively described by van den Bosch *et al.* (2008). At the One Stop Bleeding Clinic the patients underwent a grey scale ultrasound (n = 402) followed by colour Doppler examination (n = 402), followed by contrast sonohysterography (n = 398), office hysteroscopy (n = 381) and endometrial sampling (n = 243). If indicated the patients underwent operative hysteroscopy (n = 131) or hysterectomy (n = 14).

At grey scale ultrasound the total endometrial thickness was measured in the midsagittal plane. The ultrasound examiner also reported the presence or absence of an intracavitary lesion, and if applicable, the type of lesion (e.g. endometrial polyp, intracavitary myoma). At color Doppler examination the total endometrial thickness was measured again and the presence or absence of a pedicle artery (Timmerman *et al.*, 2003) was recorded. The presence or absence of an intracavitary lesion, and the type of lesion was also reported during saline infusion sonography (SIS) and office hysteroscopy. The histology results at endometrial sampling were classified as nominal variable: "abnormal" (including endometrial polyps, intracavitary myoma, endometrial hyperplasia and endometrial malignancy), "normal" (including endometrial atrophy, proliferative- and secretory changes of the endometrium) or "no histology" (in the absence of a histology result). The "outcome" to predict was the "final diagnosis". The "final diagnosis" was classified as "normal" (including endometrial atrophy, proliferative- and secretory changes of the endometrium) or "abnormal" (including endometrial polyps, intracavitary myoma, endometrial hyperplasia and endometrial malignancy). The final diagnosis was based on ultrasound with SIS, hysteroscopy, endometrial biopsy, operative hysteroscopy and hysterectomy findings in 8.0%, 16.2%, 39.8%, 32.6% and 3.5%, respectively. Sensitivity is defined as the proportion of patients with an abnormal diagnosis that are correctly identified as such, while specificity refers to normal diagnosis.

Two groups of parameters were considered in the prediction of the "final diagnosis": "pre-test parameters" (including patient's age, weight, height, parity, menopausal status; presence or absence of abnormal bleeding symptoms; the result of a cervical cytology smear within the last 6 months) and "post-test parameters" (including the total endometrial thickness as measured at grey scale ultrasound, the presence or absence of an intracavitary lesion at grey scale ultrasound, the type of intracavitary lesion seen at ultrasound, the presence or absence of a "pedicle artery sign", the endometrial thickness measured at color Doppler imaging, the presence or absence of an intracavitary lesion at SIS, the type of intracavitary lesion seen at SIS, the presence or absence of an intracavitary lesion at office hysteroscopy, the type of intracavitary lesion seen at hysteroscopy, the histology of the endometrial sampling).

The data set was split into a "training set" (first 70%: 281 patients) and a "test set" (last 30%: 121 patients). The decision trees were built on the training set. The reported performance values are presented for the test set on which the decision trees are validated. To make the test results comparable, only 94/121 test patients without missing values for any of the variables included in the studied decision trees were used for validation.

The decision trees are built using an iterative process: a decision tree is first built on the complete training set of 281 patients with and without a cross-validation strategy with 10 folds. Patients with missing values are spread over multiple branches. The important variables singled out in these trees are selected, and in the next round patients with missing values for these selected variables are excluded and a new decision tree is built on the reduced training set. Again resulting decision trees will incorporate some of the variables, while others are not used. In the next round the patients with missing values for the latest selected variables are excluded and another decision tree is built that will select some of the variables. The final decision tree obtained with cross-validation is chosen based on its highest cross-validation training performance with a small difference in full and cross-validation training performance to avoid overfitting. Furthermore, reduced-error pruning is applied to reduce the chance on overfitting the training data, which means that a large tree is grown before replacing some branches by a terminal node. The training set is split into three folds of which one is used for pruning and the rest for growing the tree. After building a decision tree on the training set, patients with missing values for one or more of the included

variables are removed from the test set before validating the decision trees.

First a decision tree has been built using both "pre-test" and "post-test" parameters. Second a decision tree only based on "post-test" parameters was built. Finally a decision tree was designed without using the hysteroscopy variables.

The Waikato Environment for Knowledge Analysis (Weka) software (version 3.4.8, University of Waikato, New Zealand) was used for the development of decision trees with the J48 algorithm, a slightly modified version of C4.5 (Quinlan, 1993).

Results

The mean patients' age was 50.7 years (SD 12.0). The average (SD) patient's weight and height was 69.9 kg (14.2) and 163.8 cm (6.1), respectively. Fifty-three percent of women were premenopausal and 12.7% were nulliparous (mean parity 1.9; SD 1.2). The mean endometrial thickness at grey scale ultrasound examination was 9.6 mm (SD 6.8). In 11 patients (2.7%) endometrial cancer was diagnosed, in 24 (6.0%) endometrial hyperplasia, in 111 (27.6) an endometrial polyp and in 48 (11.9%) an intracavitary myoma.

The first decision tree using both "pre-test" and "post-test" parameters (Tree #1) was built on the training set containing 254 patients. The selected variables were: the presence or absence of an intracavitary lesion on hysteroscopy, the parity, the menopausal status, the histology result at office endometrial sampling and the endometrial thickness as measured at grey scale ultrasound examination (Fig. 1). The test set contains 121 patients. After removing patients with missing values (for the presence or absence of an intracavitary lesion on hysteroscopy, the histology result at office endometrial sampling and the endometrial thickness at grey scale ultrasound), 101 patients are left in the test set.
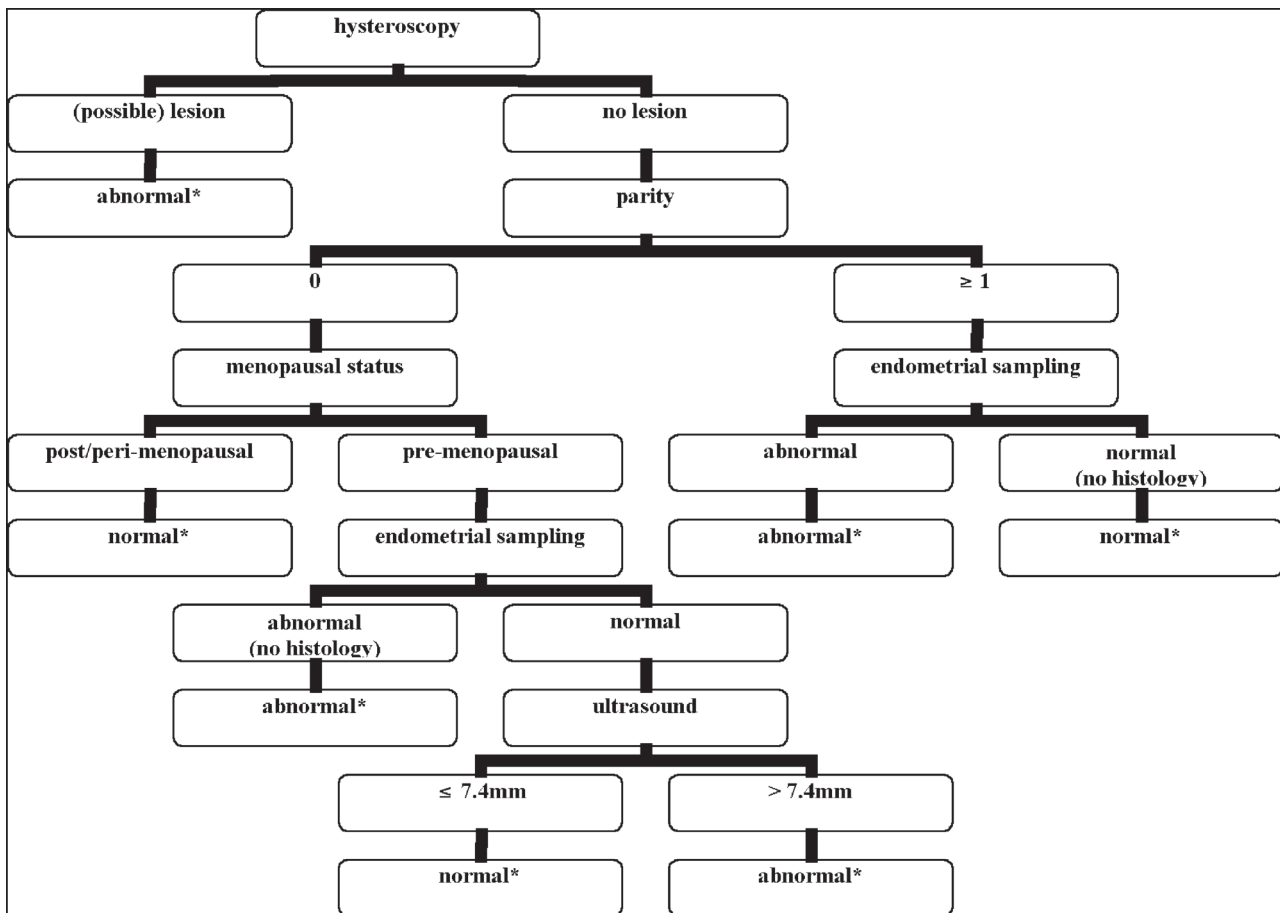

($) "pre-test parameters" include patient's age, weight, height, parity, menopausal status; bleeding symptoms; cervical cytology.
(£) "post-test parameters" include ultrasound-, color Doppler-, SIS-, hysteroscopy- findings as well as the histology results after endometrial sampling.
(*) "abnormal" was defined as the presence of benign or malignant intracavitary pathology; "normal" was defined as the absence of any intracavitary lesion, and includes endometrial atrophy as well as proliferative- and secretory endometrial changes.

Fig. 1. — Decision Tree #1 (based on both "pre-test"$ and "post-test"£ parameters)
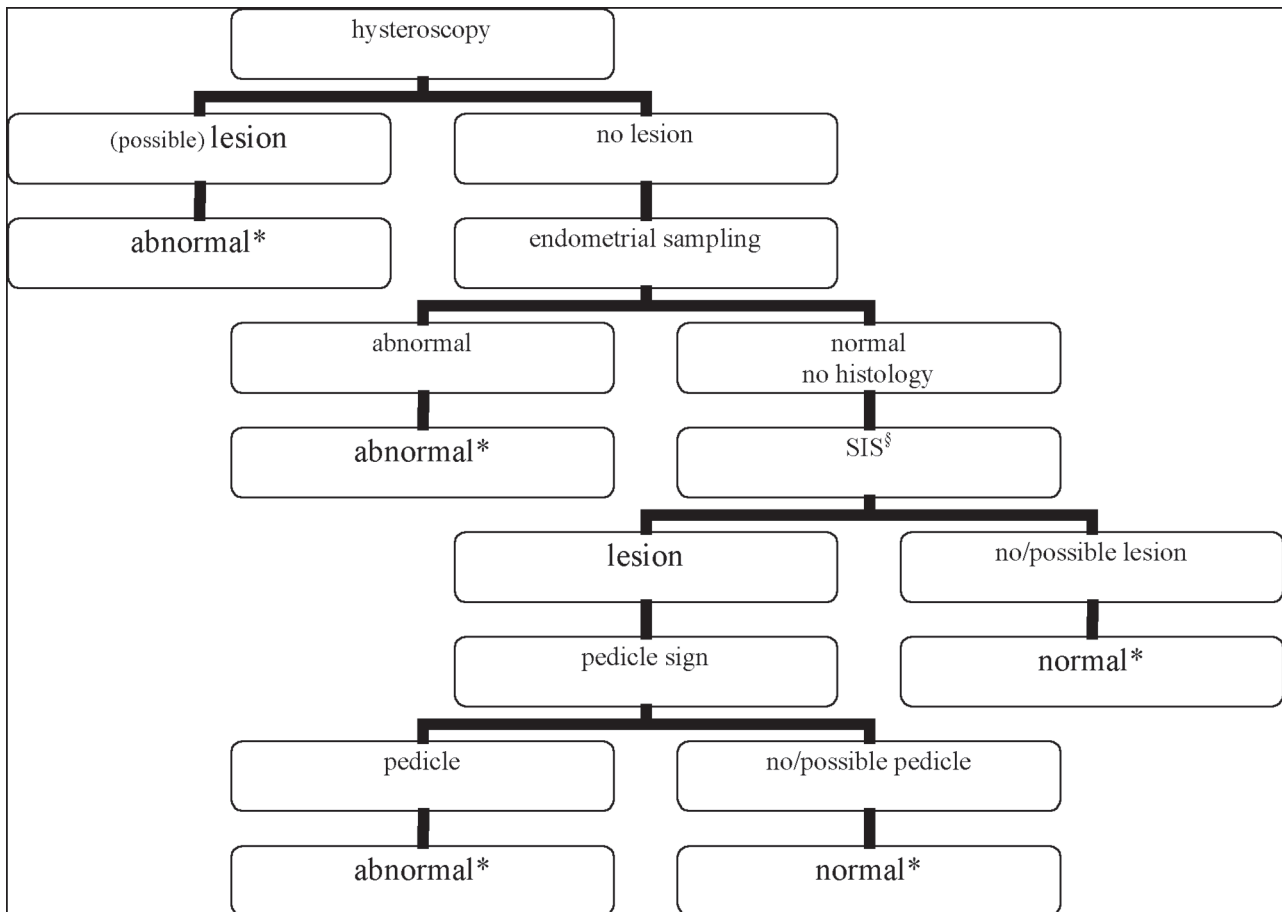
(£) "post-test parameters" include ultrasound-, color Doppler-, SIS-, hysteroscopy- findings as well as the histology results after endometrial sampling.
(§) SIS = saline infusion sonography.
(*) "abnormal" was defined as the presence of benign or malignant intracavitary pathology; "normal" was defined as the absence of any intracavitary lesion, and includes endometrial atrophy as well as proliferative- and secretory endometrial changes.
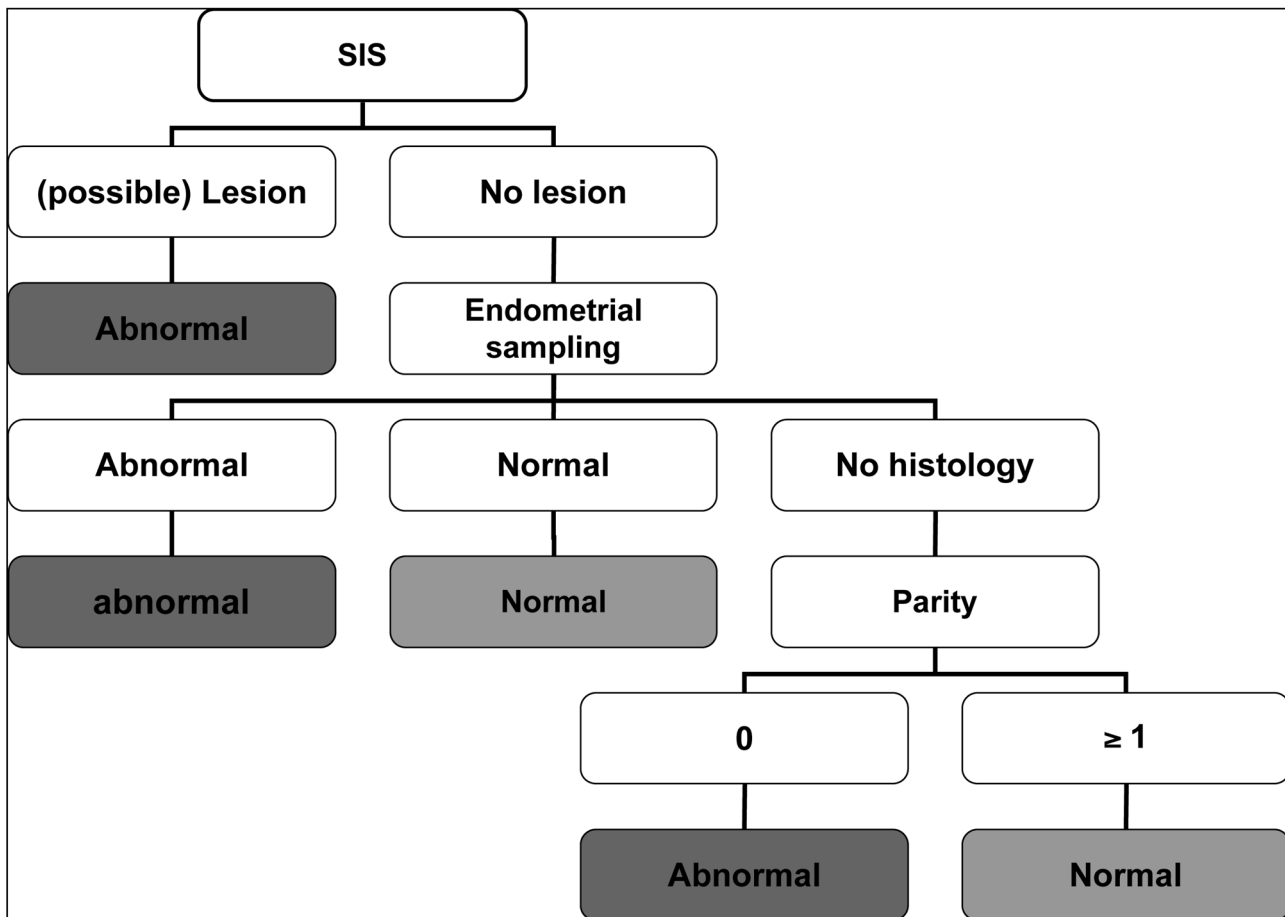
Fig. 2. — Decision Tree #2 (based only on "post-test"[£] parameters)

Ninety cases (89.1%) were correctly classified by Tree #1, the sensitivity was 95.8% and the specificity 83.0%.

The second decision tree only using "post-test" parameters (Tree #2) was built on the training set containing 243 patients. The selected variables were: the presence or absence of an intracavitary lesion on hysteroscopy, the histology results of the office endometrial sampling, the presence or absence of an intracavitary lesion on SIS and the presence or absence of a pedicle artery at color Doppler imaging (Fig. 2). After removing test patients with missing values (for the presence of an intracavitary lesion on hysteroscopy, the histology results of endometrial sampling, the presence of an intracavitary lesion on SIS), 98 patients were left in the test set. Eighty-seven cases (88.8%) were correctly classified by tree #2, the sensitivity was 97.9% and the specificity 80.4%.

The third decision tree was designed using both "pre-test" and "post-test" parameters but without using the hysteroscopy variables (Tree #3). It was built on a training set containing 268 patients. The selected variables were: the presence or absence of an intracavitary lesion on SIS, the histology results at office endometrial sampling and the patient's parity (Fig. 3). After removing test patients with missing values for the selected variables, 107 patients were left in the test set. Ninety cases (84.1%) were correctly classified by tree #3, the sensitivity was 92.3% and the specificity 76.4%.

To be able to compare the performance of the different decision trees, only those patients without missing values for any of the selected variables were included in the test set. All three decision trees were validated on a test set containing 94 patients. The diagnostic accuracy was 88.3%, 88.3% and 84.0% for tree #1, #2 and #3 respectively, the sensitivity

($) "pre-test parameters" include patient's age, weight, height, parity, menopausal status; bleeding symptoms; cervical cytology
(£) "post-test parameters" include ultrasound-, color Doppler-, SIS-, hysteroscopy- findings as well as the histology results after endometrial sampling (§) SIS = saline infusion sonography
(*) "abnormal" was defined as the presence of benign or malignant intracavitary pathology; "normal" was defined as the absence of any intracavitary lesion, and includes endometrial atrophy as well as proliferative- and secretory endometrial changes.

Fig. 3. — DecisionTree #3 (based on both "pretest"$ and "post-test"£ parameters, but excluding the hysteroscopy variables).

and specificity was 95.5% and 82%, 97.7% and 80.0, 93.2 and 76.0%, respectively (Table 1).

Discussion

Unlike the one-to-one comparison between two tests, the method used in this study enables the comparison between different decision trees containing multiple tests. Our study used the data of individual patients to build and to validate the diagnostic algorithms. The presented decision trees were built by a non-biased mathematician and are therefore not influenced by the clinician's preferences. Other studies have built decision trees based on hypothetical likelihood ratios and assumptions extracted from other series (Clark *et al*., 2006).

Basically, all 3 trees start with an imaging technique: hysteroscopy or SIS). If no lesion was seen at first evaluation, other diagnostic steps are proposed

to lower the false negative rate. Most clinical algorithms also propose an imaging technique as cornerstone examination in the diagnosis of intracavitary lesions (Van den Bosch, 2007). Imaging, beit hysteroscopy or SIS, selects who needs endometrial sampling (e.g. in case of a diffusely thickened endometrium), who should undergo operative hysteroscopy (e.g. in case of an endometrial polyp) and who does not need further testing (i.e. in case of a thin and regular endometrium). Imaging may also act as quality control during subsequent endometrial sampling procedure: e.g. if a thickened endometrium had been seen on ultrasonography or hysteroscopy, and if endometrial sampling hardly yields any tissue, the lesion most probably has been missed during the sampling.

The decision trees were built to diagnose "intra-uterine disease" including both benign and malignant lesions. The algorithm is expected to depend on

Table 1. — Accuracy, sensitivity and specificity of the decision Trees #1, #2 and #3 in the diagnosis of intracavitary pathology*.

	Decision tree		
	Tree #1	Tree #2	Tree #3
Accuracy	88.3	88.3	84.0
Sensitivity	95.5	97.7	93.2
Specificity	82.0	80.0	76.0

(Tree #1 using both "pre-test"$ and "post-test"£ parameters; Tree #2 using only "post-test" parameters; Tree #3 using "pre-test" and "post-test" parameters without hysteroscopy data).
(*) intracavitary pathology includes both benign- and malignant disease (endometrial polyps, intracavitary fibroids, endometrial hyperplasia and endometrial cancer).
($) "pre-test parameters" include patient's age, weight, length, parity, menopausal status; bleeding symptoms; cervical cytology.
(£) "post-test parameters" include ultrasound-, color Doppler-, SIS-, hysteroscopy- findings as well as the histology results after endometrial sampling.

the prevalence of the endpoint (i.e. benign and malignant intracavitary disease) in the study population. In our series, consisting mostly of perimenopausal women presenting with abnormal bleeding, the prevalence of focal intracavitary lesions, such as endometrial polyps, was relatively high, while the prevalence of cancer was low. In another population the resulting decision tree may be different.

The results may also be influenced by the choice of the reference test in the diagnosis of intracavitary lesions. Histology together with diagnostic hysteroscopy is usually considered the gold standard in the diagnosis of intracavitary lesions. However, both histology and hysteroscopy have their limitations too: e.g. a resected endometrial polyp may get lost during the processing of the specimen (Duffy *et al.*, 2003), or a sessile endometrial lesion may remain unseen at hysteroscopy. Because of the lack of any infallible gold standard, any decision tree will be prone to some bias.

Tree #2 included the "pedicle artery sign" as last step in those women with a focal lesion seen at SIS, but not on hysteroscopy: a vessel seen at color Doppler inside a focal thickening is indicative for intracavitary pathology, whereas in the absence of any color Doppler signal an artifact (e.g. a blood clot or some endometrial tissue pushed up while treading the SIS-catheter) is more probable.

The selection of parity and menopausal status was somewhat unexpected. In tree #1, if no lesion was seen at hysteroscopy, an endometrial biopsy

was proposed straight away in the parous women, whereas, in nulliparous women, endometrial sampling was restricted to premenopausal patients: nulliparous, postmenopausal women did not seem to benefit from further testing. In tree # 3, nulliparous women were considered more suspicious for pathology. The clinical correlate for this is unclear.

The endometrial thickness is used in the last step of tree #1: an endometrial thickness above 7.4 mm is considered abnormal. It must be emphasized that this cut-off value cannot be extrapolated for use as a single test outside the algorithm, but only in the very selected cases of premenopausal nulliparous women in whom hysteroscopy failed to show a lesion and in whom endometrial sampling showed a normal histology. If used as single test the cut-off value for endometrial thickness above which malignancy is to be ruled out lies between 3 and 5mm (Tabor *et al.*, 2002; Smith-Bindman *et al.*, 1998; Gupta *et al.*, 2002; Epstein & Valentin, 2004; Timmermans, 2009).

Decision tree #3 is very simple, but still has a reasonable accuracy without third or fourth line tests to lower the false negative rate. Allowing hysteroscopy to be used in second or third line examination may have improved its diagnostic accuracy.

We do not pretend that decision trees built by mathematician are superior to algorithms based on "good clinical judgement". However, the - sometimes unexpected - results of the mathematical decision trees may lead the clinician to interesting reflection as to the current clinical practice.

In practice, the choice of a diagnostic algorithm will also be influenced by other factors, such as the personal preference and skills of the clinician, the availability of the different diagnostic methods, the possibility to use the decision tree in a "one stop clinic" setting, the patient's preference for one test (Van den Bosch *et al.*, 2008) and the cost.

References

Clark TJ, Barton PM, Coomarasamy A, Gupta JK, Khan KS. Investigating postmenopausal bleeding for endometrial cancer: cost-effectiveness of initial diagnostic strategies. BJOG. 2006;113:502-10.

Clark TJ, Mann CH, Shah HM, Khan KS, Song F, Gupta JK. Accuracy of outpatient endometrial biopsy in the diagnosis of endometrial cancer: a systematic quantitative review. BJOG. 2002;109:313-21.

Clark TJ, Voit D, Gupta JK, Hyde C, Song F, Khan KS. Accuracy of hysteroscopy in the diagnosis of endometrial cancer and hyperplasia. JAMA. 2002;288:1610-21.

de Kroon C, De Bock GH, Dieben SWM, Jansen FW. Saline contrast hydrosonography in abnormal uterine bleeding: a systematic review and meta-analysis. BJOG. 2003;110:938-47.

Dijkhuizen FPHLJ, Mol BWJ, Brölmann HAM, Heintz APM. The accuracy of endometrial sampling in the diagnosis of

patients with endometrial carcinoma and hyperplasia. A meta-analysis. Cancer. 2000;89:1765-72.

Duffy S, Jackson TL, Lansdown M, Philips K, Wells M, Pollard S, Clack G, Cuzick J, Coibion M, Bianco AR. The ATAC adjuvant breast cancer trial in postmenopausal women: baseline endometrial subprotocol data. BJOG. 2003;110: 1099-1109.

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

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

Quinlan JR. Programs for machine learning. Morgan Kaufmann Publishers, Inc., 1993.

Smith-Bindman R, Kerlikowske K, Feldstein VA, Subak L, Scheidler J, Segal M, Brand R, Gracy D. Endovaginal ultrasound to exclude endometrial cancer and other endometrial abnormalities. JAMA. 1998;280:1510-7.

Tabor A, Watt HC, Wald NJ. Endometrial thickness as a test for endometrial cancer in women with postmenopausal vaginal bleeding. Obstet Gynecol. 2002;99:663-70.

Timmerman D, Verguts J, Konstantinovic ML, Moerman P, Van Schoubroeck D, Deprest J, Van Huffel S. The pedicle artery sign based on sonography with color Doppler imaging can replace second-stage tests in women with abnormal vaginal bleeding. Ultrasound Obstet Gynecol. 2003;22: 166-71.

Timmermans A. Postmenopausal bleeding: studies on the diagnostic work-up. (Thesis University of Amsterdam) Utrecht, 2009.

Van den Bosch T, Vandendael A, Van Schoubroeck D, Wranz PAB, Lombard CJ. Combining vaginal ultrasonography and office endometrial sampling in the diagnosis of endometrial disease in postmenopausal women. Obstet Gynecol. 1995;85:349-52.

Van den Bosch T. Towards an improved diagnosis of uterine pathology. (Thesis) Leuven University press, Leuven, 2007.

Van den Bosch T, Verguts J, Daemen A, Gevaert O, Domali E, Claerhout F, Vandenbroucke V, De Moor B, Deprest J, Timmerman D. Pain experienced during transvaginal ultrasound, saline contrast sonohysterography, hysteroscopy and office sampling: a comparative study. Ultrasound Obstet Gynecol. 2008;31:346-51.