Computerised morphometrical analysis in endometrial hyperplasia for the prediction of cancer development. A long term retrospective study from northern Norway

Anne Ørbo, Jan P A Baak, Inger Kleivan, Sigrun Lysne, Per S Prytz, Marc A M Broeckaert, André Slappendel, Hans J Tichelaar

Abstract

Aims—To evaluate and compare the long term prognostic value of the WHO classification and the computerised multivariate morphometrical D score in endometrial hyperplasia. To test the reproducibility of the D score in two different centres.

Methods—Histopathological WHO classification and computerised morphometrical analysis using the D score (< 0, high risk; > 1, low risk; 0–1, uncertain) in a population based study from northern Norway of archival dilatation and curettage material from 68 women with 10–20 years of follow up.

Results—Of the 68 patients included in the study, 18 developed cancer. The sensitivity and specificity of the D score (< 0 *v* > 1) were 100% and 78%, respectively, which was better than the WHO classification (89% and 60%, respectively). The negative and positive predictive values for the D score were 100% and 58% and of the WHO classification 94% and 44%, respectively. This study found a slightly higher specificity for the D score than former retrospective studies, but otherwise the results were comparable. The D score results were reproducible between the two centres (R = 0.91; slope = 0.98; intercept = 0.3).

Conclusions—D score assessment is a reproducible and more accurate predictor of outcome of endometrial hyperplasia than the WHO classification assessed by an experienced gynaecological pathologist. Routine application of the D score might reduce over and undertreatment of endometrial hyperplasia.

(*J Clin Pathol* 2000;**53**:697–703)

Institute of Medical Biology, University of Tromsø, N-9037 Tromsø, Norway
A Ørbo
I Kleivan
S Lysne
P S Prytz

Department of Pathology, Academic Hospital, Free University, de Bokelaan 1117, 108 1HV Amsterdam, The Netherlands
J P A Baak
M A M Broeckaert

Medical Center Alkmaar, Wilhelminalaan 12, 1815 JD Alkmaar, The Netherlands
A Slappendel
H J Tichelaar

Correspondence to:
Professor Ørbo
email: patano@rito.no
jbaak@mca.alkmaar.nl

Accepted for publication
19 April 2000

For many years, endometrial hyperplasia has been a diagnostic problem for pathologists.[1 2] This is understandable because the condition comprises a spectrum of histological changes from simple exaggeration of the normal proliferated state at one extreme to changes that are difficult to distinguish from carcinoma at the other end of the spectrum.[3 4] Over the past decades, many studies on different organs have made it clear that in such continuous spectral lesions normal qualitative subjective microscopical evaluations are not completely reproducible, even among experts, and that the resulting differences may be associated with important prognostic variations.[5–7]

In the past, the existence of several classification systems and the use of descriptive diagnostic terms have resulted in low interobserver and intra-observer diagnostic reproducibility for endometrial hyperplasia. Moreover, previous studies have shown that only 10–20% of endometrial hyperplasias progress to carcinoma when left untreated.[2 3] Because the treatment of choice for patients with hyperplasia has often been hysterectomy, it is likely that overtreatment of many women with endometrial hyperplasia has occurred. Conversely, other patients with endometrial hyperplasia might have received hormonal treatment or no treatment at all as a result of underdiagnosis, even though they actually should have received more active treatment. The lack of criteria that could accurately predict the disease outcome may also have been an important cause of both over and undertreatment.

The WHO classification,[2] introduced after the important publication by Kurman *et al* in 1985, has been widely accepted and has dramatically improved the diagnostic agreement in cases of endometrial hyperplasia. However, it must be admitted that the prognostic accuracy as to cancer progression or not is still not perfect. For example, even in the worst prognostic group, only 29% of cases with complex atypical hyperplasia develop cancer. Moreover, a recent study[8] found that only "atypicality or not" was reasonably well reproducible. However, cancer progression in the patients with atypicality is even lower (~ 10%) than in the group of patients with complex atypical hyperplasia, whereas in the "no atypicality" group cancer can still develop.[2] Therefore, it has to be concluded that to date no reliable and accurate method exists to discriminate in individual patients between hyperplasias with different malignant potential for cancer development.

Earlier morphometrical studies have shown that the measurement of nuclear features can predict progression to cancer.[9–12] When quantitative architectural and nuclear criteria were combined, the accuracy of prognosis prediction in endometrial hyperplasia was even better.

The latter resulted in a computerised morphometrical analysis based multivariate

prognostic formula called the D score.[3][13][14] The D score showed a higher sensitivity than qualitative microscopical characteristics (nuclear atypicality and glandular complexity) used in the WHO classification. An independent test study confirmed these results,[15] but follow up was relatively short.

To test independently the prognostic value of the morphometrical D score, a population based study from northern Norway of archival dilatation and curettage material from 68 women with 10–20 years of follow up was undertaken. We also intended to compare the results of the quantitative method with those of the WHO classification. To assess the reproducibility of the D score, the material was investigated independently in two different quantitative pathology centres, the first (Tromsø) being recently established, the other (Amsterdam/Alkmaar) with more than 15 years experience in quantitative pathology.

Patients and methods

SUBJECTS
Histological material obtained from uterine curettings with hyperplasia from northern Norway was used in our study. The material was routinely processed in different laboratories, fixed in buffered formaldehyde, embedded in paraffin wax, and standard histological sections were made. The first curettage material on which the diagnosis hyperplasia was made was used for the analyses. No repeat biopsies were examined. If more than one section was available from the curettage, the one with the subjectively most abnormal deviation was used. The inclusion criteria in the "no progression" group were: (1) an original first diagnosis of endometrial hyperplasia and (2) no cancer development. The progression group consisted of cases that also had an original diagnosis of hyperplasia, but developed histologically confirmed endometrial cancer in the follow up (in all these progression cases, the diagnosis of cancer was confirmed in a hysterectomy specimen). The follow up was minimally 10 years and maximally 20 years for the non-progression group. The mean interval between the diagnosis of hyperplasia in the curettage and cancer in the hysterectomy specimen was 6.6 months (range, 3–39). The mean age of the whole patient group was 48.2 years (range, 28–77). Clinical information was obtained from the hospital records. Treatment in general was conservative ("wait and see"), without hormonal treatment, and in principle no hysterectomy was performed unless the pathologists expressed strong concern about the presence of malignancy (this occurred in one patient who was treated with hysterectomy three months after the initial curettage diagnosis of hyperplasia). In spite of this generally accepted policy, in 10 other cases without any "alarm" signals in the pathology report, hysterectomy was performed within two years because of continuing complaints and concern by the gynaecologists because of the term "hyperplasia" used by the pathologist. The follow up data of all the patients were derived from the clinical records, but the data have been cross checked with the Norwegian Cancer Registry to ensure that none of these patients had developed endometrial cancers treated outside of our health region (in fact, this had not occurred). Four patients who had developed other gynaecological cancers and two with "secretory hyperplasia" (a phenomenon consisting of secretory endometrium with "more than usual" crowding of glands without atypia, which is not associated with progesterone treatment and is most likely a somewhat excessive form of normal secretory endometrium) were excluded. For morphometrical analysis the hyperplastic area should be minimally 1 mm² in size (see below). For this reason, 20 cases with cytonuclear atypia in one or two glands only were excluded. In addition, three cases with very poor quality sections or poor fixation were excluded. (None of the 29 cases that were excluded for technical or diagnostic reasons described above developed endometrial cancer.) In total, 50 cases were included in the no progression group and 18 cases in the cancer progression group. None of the patients was lost from the follow up.

For histological examination and quantitative analysis the slides were mixed and blinded to the investigators (clinical follow up was disclosed only after the pathological WHO re-evaluation and morphometrical analysis). For the revision of cases, we used the criteria described in the original study by Kurman *et al* (cytonuclear atypicality—including stratification—and glandular complexity). After extensive attempts by two of us (AO and JPAB), stromal invasion was not used as a criterion because, in agreement with others, we did not find it reliable or reproducible enough. Myometrial invasion did not occur in the hyperplasia curettings used for the analyses. All cancers in the uterus that occurred in the follow up were investigated independently by two experienced gynaecological pathologists and staged according to the International Federation of Gynaecology and Obstetrics.[2]

ASSESSMENT OF MORPHOMETRICAL D SCORE
Based on the original computerised morphometrical study[3], the D score was calculated as follows:

D score = $0.6229 + 0.0439 \times$ (volume percentage stroma) $- 3.9934 \times$ Ln (standard deviation shortest nuclear axis) $- 0.1592 \times$ (outer surface density glands)

where Ln stands for natural logarithm.

One haematoxylin and eosin (HE) stained section from each patient was investigated: one of us (AO) assessed the section with the subjectively most atypical area of approximately 3×3 mm (minimally 1×1 mm) (the so called measurement area). Non-hyperplastic areas were carefully excluded. For quality control, each case was reviewed using a double headed microscope by two of us (AO, JB), and in three of the 68 cases the demarcation initially drawn was changed (non-hyperplastic adjacent areas were more carefully erased to avoid any misunderstanding). Complete agreement was obtained as to the measurement

Table 1 The number of patients in each WHO subclass and the number with cancer development according to the revised assessment

	Kurman et al			Our present study		
Type	Total (n)	With cancer (n)	With cancer (%)	Total (n)	With cancer (n)	With cancer (%)
Simple hyperplasia	93	1	1	9	0	0
Complex hyperplasia	29	1	3	10	1	10
Simple atypical hyperplasia	13	1	8	13	1	7.5
Complex atypical hyperplasia	35	10	29	36	16	44
Total	170	13	8	68	18	26

The results of Kurman's study[2] are compared with the results of our present study.

area finally demarcated for morphometrical analysis.

The measurements in Tromsø were performed by a trained and experienced technician and glandular architectural and nuclear morphometrical investigations were done separately with the QPRODIT system (version 6.1; Leica, Cambridge, UK), as described previously.[3] Because of the modern quantitative equipment now available, some technical modifications have been included that make the method faster and more reproducible, as follows. The architectural investigation was performed by a 16 point electronic grid system of eight "Weibel-type" test lines with a ×20 objective (grid point distance at specimen level was 40.5 μm) (when another camera is used, the difference in the camera magnification factor sometimes requires a ×40 objective to achieve that same grid point distance). The intersections of the grid lines with the outer surface of glands (for the assessment of outer surface density) and of the end points of the grid lines falling on the stroma and other tissue compartments (for the assessment of volume percentage stroma) were registered. Points falling on glandular lumina were counted as such, but other empty spaces (such as scars, no tissue) were ignored. When all points falling on stroma, glandular epithelium, and luminar spaces were counted, and also the intersections of the grid lines with the inner and outer surfaces of the glands were assessed in one field of vision, the next field of vision was selected using a random systematic sampling approach in the measurement area (see below). In each measurement section, a total of approximately 300 grid line end points falling on stroma or other tissue compartments was analysed (equalling at least 18 fields of vision). With the built in stereological formulas, the QPRODIT system calculated automatically after these measurements in each case the outer surface density of the glands and the volume percentage of stroma.

The measurements of the shortest nuclear axis were performed in the measurement area with a ×100 (numerical aperture, 1.30) oil immersion objective, and at least 150 nuclei were systemically randomly selected using an automated scanning stage controlled by the QPRODIT system, as follows. At least 100 fields of vision were randomly equidistantly placed by the QPRODIT system over the whole measurement area. In each field of vision, an 8.1 μm (at specimen level) equidistant points grid was superimposed, and the first three clearly visible glandular nuclei hit by a grid point (meandering from top left to bottom right) were measured. The shortest nuclear axis of the nuclear profile was measured with the straight line length module of QPRODIT. If adequate glandular nuclei were present in the field of vision but none of these nuclei was hit by a grid point, the stage was moved to the next field of vision and the procedure was repeated. In this way, the selection of nuclei was truly random both within each field of vision and also in the whole measurement area. Although this form of point sampling preferentially selects larger nuclei, it gives very reproducible results between observers,[15 16] and does not influence the diagnostic value of the procedure. After each nucleus measured, the QPRODIT system calculated the coefficient of error (CE) of the running mean, and from 150 nuclei onwards, measurements were terminated if the CE dropped below 5%. Using this method, D score assessment can be done by an experienced technician within 30 minutes and thus is clinically feasible.

TEST OF REPRODUCIBILITY

To test the reproducibility and robustness of the D score, morphometrical investigations of the material were performed in Tromsø, Norway, and in Alkmaar, the Netherlands, according to the same protocol.

STATISTICAL ANALYSIS

The sensitivity, specificity, and negative and positive predictive value of the WHO classification and the D score to predict cancer progression were calculated.[17] The reproducibility of the D score between Tromsø and Alkmaar was assessed by calculating the linear correlation coefficient, the slope, and the intercept of the D score function with the y-axis. In addition, confusion tables were made for the D scores (with the decision thresholds D score $<0v>0$, $<1v>1$, and $<v>1$), the WHO classification, and "progression or not".

Results

Of the 68 patients included in our study, 18 developed cancer (table 1). When the cancers were staged according to the FIGO system, most had superficial disease: 10 patients had stage Ia, seven had Ib, and only one had FIGO Ic endometrial cancer.

PATHOLOGISTS ASSESSMENT

Table 1 shows how the 68 cases were classified according to the WHO classification in relation to the follow up. The percentage of simple hyperplasias in our study was considerably lower than in Kurman's original publication, but the percentage distribution of the other three WHO subgroups was comparable with former studies.[2 13 15] This suggests a more conservative diagnostic approach in the past in Tromsø in making the diagnosis of simple hyperplasia compared with Kurman's policy, but a comparable diagnostic attitude to the subclassification of non-simple hyperplasias. Thus, it is likely that many cases that would have been classified as simple hyperplasias by Kurman were classified in the past in Tromsø

Table 2 The number of patients with and without cancer in the different groups of D score

	D score <0	D score 0–1	D score >1	Total
No cancer	8	13	29	50
Cancer	11	7	0	18
Total	19	20	29	68
% With cancer	58%	35%	0%	26%

as no hyperplasia, and therefore were not included in our study. This also explains the higher cancer incidence (26.6%) in our study compared with Kurman's publication (8%).

Of the 49 patients who were characterised as having atypical hyperplasia, 17 progressed to cancer and many had complex atypical hyperplasia (n = 36). Complex atypical hyperplasia versus no complex atypical hyperplasias had a sensitivity of 89%, a specificity of 60%, and negative and positive predictive values of 94% and 44%, respectively, for cancer development. For "atypicality or not", sensitivity was 35%, specificity was 95%, and negative and positive predictive values were 36% and 94%, respectively. These values are comparable with Kurman's publication[2] and earlier morphometrical studies[13 15] (table 1).

D SCORES

The D score results were reproducible between the two centres (R = 0.91; p < 0.0001; slope = 0.98; intercept = 0.3). The D score distribution is shown in table 2. Of the 19 patients with a D score < 0, 11 developed cancer. Of the 20 cases with a D score > 0 < 1, seven developed cancer during follow up. The group with D score > 1 comprised 29 patients, none of whom developed endometrial cancer. All 18 patients with cancer had a hysterectomy. Eleven other patients who underwent hysterectomy had no cancer and only three of these 11 women had a D score < 0 in the original curettage (as mentioned above, in only one of these 11 patients the pathologist expressed concern as to cancer in the curettage histopathology report, in the other 10 the term hyperplasia was used but no other alarming words).

As many as 20 of the 68 patients had D score values in Tromsø between 0 and 1, for which the prognosis in previous studies

was regarded as uncertain. Interestingly, several of these Tromsø D score values were just > 0 or just < 1. The D score reproducibility assessments performed in the pathology laboratory in the Netherlands with more experience in quantitative pathology showed that far fewer (eight instead of 20) of these patients were classified into this group and consequently had values < 0 or just > 1 (in agreement with previous studies). This might be because of lack of practice with morphometrical assessments in Tromsø. It certainly might have been relevant that the old tissue blocks in this retrospective study were not optimal. Nevertheless, also in Tromsø, on average, the sensitivity, specificity, and positive and negative predictive values of the D score were better than those of the WHO classification (table 3).

RELATION BETWEEN WHO CLASSIFICATION AND D SCORE

Twelve of the 36 patients diagnosed as having complex atypical hyperplasia had a D score > 1 (table 3), and none of these developed cancer. Moreover, six of the nine patients with complex atypical hyperplasia and a D score between 0 and 1 had no cancer. Of the 15 patients with complex atypical hyperplasia and a D score < 0, 10 developed cancer. Thus, in these patients the D score clearly gave strong prognostic information. Seven of the patients with simple atypical hyperplasia had a D score > 1 and none developed cancer. One of the six patients with simple atypical hyperplasia and a D score < 1 developed cancer (D score was 0.1 in this patient).

One patient with simple hyperplasia and two with complex hyperplasia had a D score < 0. The patient with simple hyperplasia died at 24 months follow up from cerebral haemorrhage. One of the patients with complex hyperplasia had a D score of −0.2 and after 204 months follow up had not developed cancer. The other patient with complex hyperplasia had a D score of −1.2 and was cancer free, but follow up was only 12 months. This patient had undergone hysterectomy and the diagnosis based on the hysterectomy was adenomatous hyperplasia without atypia (complex hyperplasia).

Table 3 The distribution of the diagnostic assessments of the pathologist related to the different groups of the D score and selected patient details

	D score <0	D score 0–1	D score >1	Total	
Diagnostic assessment related to D score					
SH	1 (case 1)	2 (cases 2,3)	6	9	
CH	2 (cases 4,5)	4 (cases 6–9)	4	10	
SAH	1	5	7	13	
CAH	15	9	12	36	
Total	19	20	29	68	
Patient details					
Case no	*D score*	*WHO*	*Hysterectomy*	*Follow up (months)*	
Case 10	−0.4	SH	No	Died from cerebral haemorrhage (24)	
Case 40	0.6	SH	No	Alive and well (180)	
Case 6	0.9	SH	No	Alive and well (204)	
Case 66	−0.2	CH	No	Alive and well (204)	
Case 1	−1.2	CH	Yes (CH)	Hysterectomy (12)	
Case 13	0.1	CH	No	Alive and well (92)	
Case 49	0.6	CH	No	Alive and well (180)	
Case 9	0.3	CH	Yes (CH)	(1)	
Case 76	1.0	CH	Yes (CH)	(1)	

CAH, complex atypical hyperplasia; CH, complex hyperplasia; SAH, simple atypical hyperplasia; SH, simple hyperplasia.

COMPARISON OF THE DIFFERENT METHODS

Table 4 summarises the sensitivity, specificity, and negative and positive predictive values of different predictive methods. The sensitivity of the D score (either as < 0, > 1 or as < 1, > 1) is 100%, which is better than the WHO classification. Specificity is also highest with these D score thresholds. Only "complex atypical hyperplasia or not" has a similar specificity, positive predictive value, and overall correctly classified proportion of cases as D score < 1, > 1, but the sensitivity and negative predictive values of complex atypical hyperplasia are lower than the D score. The sensitivity, specificity, and negative and positive predictive values for the D score used in previous studies (< 0, > 1) is highest—100%, 78%, 100%, and 58%. In our present study, a slightly higher specificity for the D score has been obtained

Table 4   Sensitivity and specificity of the diagnostic assessment of the pathologists and the morphometrical D score

	Sensitivity (%)	Specificity (%)	NPV (%)	PPV (%)	Percentage correct
Atypicality (CAH + SAH) or not	94	36	95	35	51
Complex atypical hyperplasia (CAH) or not	89	60	94	44	68
Complexity and atypicality (CH + SAH + CAH) or not	100	18	100	31	40
D score <0, >1	100	78	100	58	83
D score <0, >0	61	84	86	58	78
D score <1, >1	100	58	100	46	69

CAH, complex atypical hyperplasia; CH, complex hyperplasia; NPV, negative predictive value; PPV, positive predictive value; SAH, simple atypical hyperplasia.

than in one former study from our group,[13] although it is comparable with the results of Dunton *et al.*[15]

Discussion

The diagnosis of hyperplasia in the past has often led to hysterectomy, even though only a relatively small proportion of cases with endometrial hyperplasia is associated with cancer in the follow up.[2 13 15] In our present study, a total of 26% of the patients developed malignancy, a figure close to that in two other studies.[14 15] The percentage distribution of patients with cellular atypicality and glandular complexity was also comparable with those studies, but in Kurman's study many more simple hyperplasias occurred. It is likely that many cases that would have been classified as simple hyperplasias by Kurman were classified in the past in Tromsø as no hyperplasia. Perhaps they have been classified as so called "disorderly proliferation" lesions and thus were not included in our present study. This explains the lower cancer incidence in Kuman's study.

The WHO classification for endometrial hyperplasia[2] intended to improve reproducibility, diagnostic value, and prognostic accuracy and has proved to be a useful, widely used, diagnostic classification system. However, prognosis prediction is still not as accurate as is desired. This is because the reproducibility of subjective grading systems might not always be perfect.[5 6] The computerised morphometrical D score in our study was found to be reproducible, to have strong prognostic value, and to take only 30 minutes of a technician's time in a routine setting. Quantitative pathological analysis can thus be of help in routine pathology practice; its advantages are objectivity and the possibility of detecting differences and changes that otherwise would escape human observation.[6] The possibility of giving more accurate, prognosis based treatment advice to clinicians fits in with the recent trend of evidence based medicine.[18] As to the accuracy of prognosis prediction of the D score, in former studies based exclusively on nuclear morphometry,[9 10] the outcome of approximately 15% of the patients with non-progressive disease could be predicted accurately.[9 11] However, the combination of quantitative nuclear and architectural criteria in the prognostic morphometrical D score gave better results and improved the prognostic power of computerised morphometrical analysis compared with nuclear morphometry

alone in estimating the outcome of hyperplasias.[3]

Comparison of the WHO classification, clinical outcome, and the D score showed that of the 36 patients diagnosed as having complex atypical hyperplasia, 20 did not develop cancer with a long follow up, and 18 of these 20 women had a D score > 0. Likewise, of the 11 women with simple atypical hyperplasia and a D score > 0 none developed cancer. Of the 11 patients with simple hyperplasia who underwent hysterectomy but cancer was not found in the hysterectomy specimen, only three had a D score < 0 in the preceding curettage (interestingly, in only one of these cases the original pathology report gave an "alarm" that the endometrial hyperplasia was atypical). This means that eight of these 11 patients would not have undergone major surgery if the treatment decision had been based on D score criteria set out in our previous studies. In our present study, none of the 29 patients with endometrial hyperplasia with a D score > 1 developed cancer, whatever the WHO diagnosis was. Consequently, with the therapeutic decision rule D score > 1.0, 29 of the 68 cases could have been correctly predicted as having a very low probability for cancer development. Thus, in these patients, hysterectomy could have been avoided. For patients with simple and complex hyperplasia, an even lower D score value threshold (> 0, which occurred in six cases, in addition to the 10 with a D score > 1) seems an appropriate and safe way to indicate a very low probability of cancer development. Using this rule, in our present study the D score clearly would have prevented overtreatment in 45 patients with endometrial hyperplasia without causing undertreatment in the 18 cases with cancer progression.

Thus, the D score has stronger prognostic value than the WHO classification and is very reproducible between independent centres. This could have therapeutic consequences, as long as the pathologist in charge carefully demarcates the measurement area and also checks the reproducibility of measurements. Clinical usefulness is also guaranteed because D score assessment can be done by a trained technician using fairly simple equipment and the method only requires a good quality standard HE histological section. Thus, computerised morphometrical analysis of endometrial hyperplasia sections is inexpensive and generally applicable; even when this technique is unavailable (such as in small laboratories), the

HE sections could still be sent to a bigger laboratory for assessment.

In spite of the strength of the computerised morphometrical D score assessment, the accuracy of the method still needs to be refined. The large group of cases with a D score between 0 and 1 may partly be a result of the sometimes low quality of the oldest sections. The nuclei in some of these slides were difficult to separate. Overlapping nuclei regarded as one might result in higher values of the shortest nuclear axis, either higher or lower SD values, and consequently a lower or higher D score. However, the occurrence of cases with a D score > 0 but < 1 is not an essential prognostic problem, because repeated measurements can indicate the "real" D score value. Moreover, even if the D score value remains uncertain, our current policy is to perform re-curettage after four weeks of progesterone ablation treatment. Thus, none of these cases would have been undertreated.

A relatively weak point of the morphometrical D score method is the tissue sampling at curettage. It has been found that even trained gynaecologists sometimes obtain only about 40% of the total endometrial mucosa when doing a dilatation and curettage.[19] Curettage sampling error is therefore one of the possible pitfalls of the D score morphometrical prediction method. However, underdiagnosis did not occur in the present and previous studies. Moreover, with modern methods of hysteroscopy, sampling error can be minimised because biopsies can be taken non-randomly under direct visualisation of the endometrium.

An important consideration of routine application of the D score is the favourable cost benefit analysis. For the Netherlands, we have calculated that nationwide application (consisting of one morphometry system + microscope written off in five years + 50% technician) would cost approximately 1.8 million ECU/year (1.8 million U$) (50% technician is a high estimate of what is needed in working time). The prevention of 20% (low estimate) of unnecessary hysterectomies in endometrial hyperplasia would save approximately 10 million ECU/year in hospital costs alone (loss of working days and other costs for the patient not included). Thus, the financial benefit of nationwide application of the D score is favourable. The good reproducibility of the D score and the prevention of unnecessary suffering for the patient are even more important arguments.

In spite of all the abovementioned positive points of computerised morphometrical analysis on endometrial hyperplasia, the accuracy is not 100%. In the search for other objective methods for predicting cancer development in endometrial hyperplasia, the value of DNA flow cytometrical analysis has been investigated, but this method was not found to be useful in distinguishing between endometrial hyperplasia with and without cancer development.[20 21] Recently published data have shown that some endometrial hyperplasias are monoclonal when tested for replication errors or X-chromosome inactivation.[22–24] Normal endometrium is polyclonal, but endometrial polyps can also be monoclonal.[25] Thus, monoclonality of endometrial hyperplasia is not synonymous with cancer. Monoclonal endometrial hyperplasias present morphologically with quite different patterns. In a recent study, the D score was found to distinguish between monoclonal and polyclonal endometrial hyperplasias, although several cases with a low D score were polyclonal. This may have been because of contamination with polyclonal stromal cells.[25]

Conclusions

The D score assessment is very reproducible and a more accurate predictor of outcome of endometrial hyperplasia than the WHO classification assessed by an experienced gynaecological pathologist. Using the D score, individual treatment recommendations may be given to clinicians and over and undertreatment should be reduced.

1 Winkler B, Alvarez S, Richart RM, *et al.* Pitfalls in the diagnosis of endometrial neoplasia. *Obstet Gynecol* 1984;**64**:185–94.
2 Kurman RJ, Kaminski PF, Norris HJ. The behavior of endometrial hyperplasia. A long-term study of "untreated" hyperplasia in 170 patients. *Cancer* 1985;**56**:403–12.
3 Baak JPA, Nauta JJP, Wisse-Brekelmans ECM, *et al.* Architectural and nuclear morphometrical features together are more important prognosticators in endometrial hyperplasia than nuclear features alone. *J Pathol* 1988:**154**; 335–41.
4 Ferenczy A, Gelfand MM, Tzipris F. The cytodynamics of endometrial hyperplasia and carcinoma. A review. *Ann Pathol* 1983;**3**:189–202.
5 Baak JPA, Oort J. The case for quantitative pathology in cancer diagnosis and treatment. In: *Manual of quantitative pathology in cancer diagnosis and prognosis*. New York: Springer, 1991:7–18.
6 Langley FA, Baak JPA, Oort J. Diagnosis: error sources. In: *Manual of quantitative pathology in cancer diagnosis and prognosis*, New York: Springer, 1991:19–27.
7 Baak JPA, Langley FA, Talerman A, *et al.* The prognostic variability of ovarian tumor grading by different pathologists. *Gynecol Oncol* 1987;**27**:166–72.
8 Bergeron C, Nogales FF, Masseroli M, *et al.* A multicentric European study testing the reproducibility of the WHO classification of endometrial hyperplasia with a proposal of a simplified working classification for biopsy and curettage specimens. *Am J Surg Pathol* 1999;**23**:1102–8.
9 Colgan TJ, Norris HJ, Foster W, *et al.* Predicting the outcome of endometrial hyperplasia by quantitative analysis of nuclear features using a linear discriminant function. *Int J Gynecol Pathol* 1983;**1**:347–52.
10 Ausems EWMA, Kamp JK van der, Baak JPA. Nuclear morphometry in the determination of the prognosis of endometrial hyperplasia. *Int J Gynecol Pathol* 1985;**4**:180–5.
11 Baak JPA. Further evaluation of the practical applicability of nuclear morphometry for the prediction of the outcome of atypical endometrial hyperplasia. *Anal Quant Cytol Histol* 1986;**8**:46–8.
12 Norris H J, Becker RL, Mikel UV. A comparative morphometric and cytophotometric study of endometrial hyperplasia, atypical hyperplasia, and endometrial carcinoma. *Hum Pathol* 1989;**20**:219–23.
13 Baak JPA, Wisse-Brekelmans, Fleege JC, *et al.* Assessment of risk of endometrial cancer in hyperplasia, by means of morphological and morphometrical features. *Pathol Res Pract* 1992;**188**:856–9.
14 Baak JPA. The role of computerized morphometric and cytometric feature analysis in endometrial hyperplasia and cancer prognosis. *J Cell Biochem* 1995;**23**(suppl): 137–46.
15 Dunton CJ, Baak JPA, Palazzo JP, *et al.* Use of computerized morphometric analyses of endometrial hyperplasias in the prediction of coexistent cancer. *Am J Obstet Gynecol* 1996; **174**:1518–21.
16 Baak JPA, Ladekarl M, Sørensen FB. Reproducibility of mean nuclear volume and correlation with mean nuclear area in breast cancer. *Hum Pathol* 1994;**25**:80–5.
17 Galen RS, Gambino SR. *Beyond normality: the predictive value and efficiency of medical diagnoses.* New York: Wiley, 1975.

18 Fitch MI, Gray RE, Covens A, *et al*. Gynecologists' perspectives regarding ovarian cancer. *Cancer Prev Control* 1999;**3**: 68–76.

19 Grimes DA. Diagnostic dilatation and currettage: a reappraisal. *Am J Obstet Gynecol* 1982;**142**:1–6.

20 Micael H, Kotylo PK, Mohr M, *et al*. DNA ploidy, cell cycle kinetics, and low versus high grade atypia in endometrial hyperplasia. *Am J Clin Pathol* 1996;**106**:22–8.

21 Baak JPA, Kuik DJ, Bezemer PD. The additional prognostic value of morphometric nuclear arrangement and DNA-ploidy to other morphometric and stereologic features in endometrial hyperplasias. *International Journal of Gynecological Cancer* 1994;**4**:289–97.

22 Mutter GL, Boynton KA, Faquin WC, *et al*. Allelotype mapping of unstable microsattelites establishes direct lineage continuity between endometrial precancers and cancer. *Cancer Res* 1996;**56**:4483–6.

23 Mutter GL, Boynton KA. X chromosome inactivation in the normal female genital tract: implications for the identification of neoplasia. *Cancer Res* 1995;**55**:5080–4.

24 Jovanovic AS, Boynton KA, Mutter GL. Uteri of women with endometrial carcinoma contain a histopathologic spectrum of monoclonal putative precancers, some with microsattelite instability. *Cancer Res* 1996;**56**:1917–21.

25 Mutter GL, Baak JPA, Crum PC, *et al*. Endometrial precancer diagnosis by histopathology, clonal analysis and computerized morphometry. *J Pathol* 2000;**19**:462–9.