

Reproducibility of current classifications of endometrial endometrioid glandular proliferations: further evidence supporting a simplified classification

Jaume Ordi, Christine Bergeron,¹ David Hardisson,² W. Glenn McCluggage,³ Harry Hollema,⁴ Ana Felix,⁵ Robert A Soslow,⁶ Esther Oliva,⁷ Fattaneh A Tavassoli,⁸ Isabel Alvarado-Cabrero,⁹ Michael Wells¹⁰ & Francisco F Nogales¹¹

Department of Pathology, CRESIB (Centre de Recerca en Salut Internacional de Barcelona), Hospital Clinic, Faculty of Medicine, University of Barcelona, Barcelona, Spain, ¹Laboratoire CERBA, St Ouen L'Aumône, France, ²Department of Pathology, IdiPAZ, Hospital La Paz Institute for Health Research, Autonomous University of Madrid, Madrid, Spain, ³Department of Pathology, Royal Group of Hospitals, Belfast, UK, ⁴Department of Pathology, University of Groningen, Groningen, The Netherlands, ⁵Department of Pathology, Institute of Oncology, Lisbon, Portugal, ⁶Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA, ⁷Department of Pathology, Massachusetts General Hospital, Boston, MA, USA, ⁸Department of Pathology, Yale University School of Medicine, New Haven, CT, USA, ⁹Department of Pathology, Hospital Star Medica, Santa Fe, Mexico, ¹⁰Department of Oncology, University of Sheffield, Sheffield, UK, and ¹¹Department of Pathology, University of Granada, Granada, Spain

Date of submission 3 July 2013

Accepted for publication 7 August 2013

Published online Article Accepted 12 August 2013

Ordi J, Bergeron C, Hardisson D, McCluggage W G, Hollema H, Felix A, Soslow R A, Oliva E, Tavassoli F A, Alvarado-Cabrero I, Wells M & Nogales F F

(2013) *Histopathology*

Reproducibility of current classifications of endometrial endometrioid glandular proliferations: further evidence supporting a simplified classification

Aims: To compare the reproducibility of the current (2003) World Health Organization (WHO), endometrial intraepithelial neoplasia (EIN) and European Working Group (EWG) classifications of endometrial endometrioid proliferations.

Methods and results: Nine expert gynaecological pathologists from Europe and North America reviewed 198 endometrial biopsy/curettage specimens originally diagnosed as low-grade lesions. All observers were asked to classify the cases by using the categories described in each scheme: six for WHO, four for EIN, and three for EWG. The results were evaluated by kappa statistics for more than two observations. The analysis was repeated using only two major categories (benign versus atypical/carcinoma). Both the WHO and EIN classifications showed poor interobserver

agreement ($\kappa = 0.337$ and $\kappa = 0.419$, respectively), whereas the EWG classification showed moderate agreement ($\kappa = 0.530$). Full agreement between pathologists occurred in only 28% for the WHO classification, 39% for the EIN classification, and 59% for the EWG classification. With only two diagnostic categories, kappa values increased in all classifications, but only the EWG classification reached a substantial level of agreement ($\kappa = 0.621$); similarly, full agreement among all pathologists increased to 70% for the WHO classification, 69% for the EIN classification, and 72% for the EWG classification.

Conclusions: A two-tier classification of endometrial endometrioid proliferative lesions improves reproducibility, and should be considered for the diagnosis of endometrial biopsy/curettage specimens.

Keywords: endometrial carcinoma, endometrial hyperplasia, interobserver variability

Address for correspondence: Professor F F Nogales, Department of Pathology, Facultad de Medicina, University of Granada, Av de Madrid 11, 18012 Granada, Spain. e-mail: fnogales@ugr.es

Introduction

Endometrial biopsy and curettage are essential procedures in the evaluation of abnormal vaginal bleeding, the most common presenting symptom of endometrial carcinoma and its precursors.^{1,2} Accurate diagnosis of endometrial samples should guide appropriate treatment of carcinomas and high-risk precursor lesions, while avoiding overtreatment of benign or non-progressive proliferations. However, not infrequently, the endometrium shows a continuum of abnormal glandular proliferations, ranging from benign disordered to architecturally crowded carcinoma precursors, and finally to endometrioid carcinoma.^{3,4} This range of changes, with frequent morphological overlap, results in marked difficulties in interpretation of the histological findings and their subsequent classification.

As a consequence, several classification schemes for carcinoma precursors have been introduced over the years, but there is no general consensus among pathologists about the most accurate system and nomenclature to be used. Thus, different groups have adopted and used different classifications in routine practice. The World Health Organization (WHO) system, the oldest and most widely used classification, separates endometrial proliferations into simple or complex hyperplasia on the basis of architectural features, and typical or atypical on the basis of cytological features, as originally defined by Kurman *et al.* in 1985.^{3,4} This terminology was adopted by the WHO in 1994, because of a reported increased risk of progression to endometrioid carcinoma in lesions classified as complex hyperplasia with atypia as compared with those diagnosed as complex hyperplasia without atypia.³ Previous studies evaluating the diagnostic reproducibility of the 1994 WHO system (whose categories remained unchanged in the 2003 WHO system) reported kappa values ranging from 0.2 to 0.7 for overall interobserver agreement in diagnosing endometrial hyperplasia.^{5–8}

In 1999, Bergeron *et al.*,⁵ in a study from a group of European gynaecological pathologists, proposed a simplified working classification to overcome the poor reproducibility of the WHO system. The new classification (European Working Group, EWG), which was intended to be used only on biopsy specimens, showed good reproducibility and had two major diagnostic categories: (i) 'hyperplasia', including simple and complex non-atypical hyperplasia; and (ii) 'endometrioid neoplasia', which included atypical hyperplasia and low-grade endometrioid carcinoma.

In 2000, another group of pathologists proposed a new classification to encompass emergent molecular,

histomorphometric and clinical outcome data studies. The authors introduced the term endometrial intraepithelial neoplasia (EIN) to define a premalignant lesion clonally related to invasive endometrioid carcinoma, and proposed diagnostic morphological criteria for these premalignant lesions.⁹ They include architectural thresholds (gland area exceeds that of stroma), cytological changes relative to background normal glands, and a minimum lesion size (1 mm) within a single tissue fragment. The authors found this classification to have better diagnostic reproducibility than the 2003 WHO system.⁴ Within the last few years, a considerable number of studies have reported an apparent advantage of the EIN over the WHO classification. However, reproducibility studies in support of the EIN concept were designed and performed by collaborating, but not independent, research groups.^{8–14}

Although the reproducibility of these classifications has been evaluated in a number of reports, all of these studies have evaluated a single system or compared a newly proposed classification with the WHO system; no study has compared the reproducibility of all three classification systems. Thus, our objective was to compare the interobserver reproducibility of the three currently reported classification systems for endometrial hyperplasia – WHO, EIN, and EWG – in order to provide some insights into potential strategies to improve future classifications.

Materials and methods

CASE SELECTION

The specimens comprised endometrial biopsies and curettages (henceforth referred to as 'biopsies' for simplicity) obtained from the files of three gynaecological pathologists who designed the study (J.O., D.H., and F.F.N.). Eligibility requirements included cases initially diagnosed as low-grade endometrioid lesions characterized by focally or diffusely distributed crowded glands, which included 39 lesions involving endometrial polyps. Consensus by a three-member panel (J.O., D.H., and F.F.N.) was used to select one slide that represented the most crowded glandular proliferation. A total of 198 samples were selected for the study. The samples were randomly numbered from 1 to 198. The 198 slides were divided into three equal subsets containing one-third of the samples (66 slides each).

SAMPLE EVALUATION

Histological evaluation was independently performed by nine expert gynaecological pathologists from both

Europe (C.B., M.W., H.H., A.F., and W.G.M.) and North America (E.O., R.A.S., I.A.-C., and F.A.T.). Each pathologist was assigned a random number from 1 to 9. According to the study design shown in Figure 1, observers 1–3 used method A (WHO classification) for the first third of the samples (block 1), method B (EIN classification) for the second third (block 2), and method C (EWG) for the last third (block 3). Methods for the first, second and third blocks of slides were B–C–A for observers 4–6 and C–A–B for observers 7–9. Thus, each sample was evaluated by three observers for each method. All observers were blind to any clinicopathological data other than the slide itself. Referential papers or websites concerning each classification were available to each reviewer.

All observers were asked to classify each slide by using the categories described for each classification.^{3,5,9} These included six categories for the WHO classification (cycling, atrophic, or other benign, hereafter referred to as benign; simple hyperplasia without atypia; complex hyperplasia without atypia; simple hyperplasia with atypia; complex hyperplasia with atypia; and carcinoma),³ four categories for the EIN classification (benign; benign hyperplasia; EIN; and carcinoma),⁹ and three significant categories for the

EWG classification (benign; hyperplasia; and endometrioid neoplasia).⁵

Two slides were broken during transport, and consequently excluded from the analysis, resulting in 196 adequate specimens for study.

STATISTICAL ANALYSIS

Kappa statistics for more than two observations and more than two observers (but a constant number of them) were computed for each method using the Fleiss, Nee and Landis approximation.¹⁵ The measure calculates the degree of agreement in classification over that which would be expected by chance, and is scored as a number between 0 and 1. The strength of agreement of kappa values is, following the Landis-defined categories, as follows: 0, none beyond chance; 0–0.20, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; and 0.81–1.00, almost perfect.¹⁵ Ninety-five per cent confidence intervals (95% CIs) of the kappa statistics were estimated by bootstrap analysis with 5000 replications.¹⁶

Statistical evaluations were initially performed using the diagnostic categories described for each classification (six for WHO, four for EIN, and three for EWG). The analysis was repeated after reducing the WHO categories to four categories [1, benign; 2, hyperplasia without atypia (simple and complex); 3, atypical hyperplasia (simple and complex); and 4, carcinoma], then to three categories (1, cycling endometrium; 2, hyperplasia without atypia; 3, atypical hyperplasia and carcinoma), and finally to two categories (1, benign and hyperplasia without atypia; 2, atypical hyperplasia and carcinoma). Similarly, the analysis for the EIN classification was repeated after reducing the diagnostic categories to three (1, benign; 2, benign hyperplasia; 3, EIN and carcinoma) and eventually to two (1, benign and benign hyperplasia; 2, EIN and carcinoma). Finally, the analysis for the EWG classification was repeated after reducing the diagnostic categories to two (1, benign and hyperplasia; 2, neoplasia).

The analysis was performed using the statistical software STATA, version 12.0 (StataCorp., College Station, TX, USA).

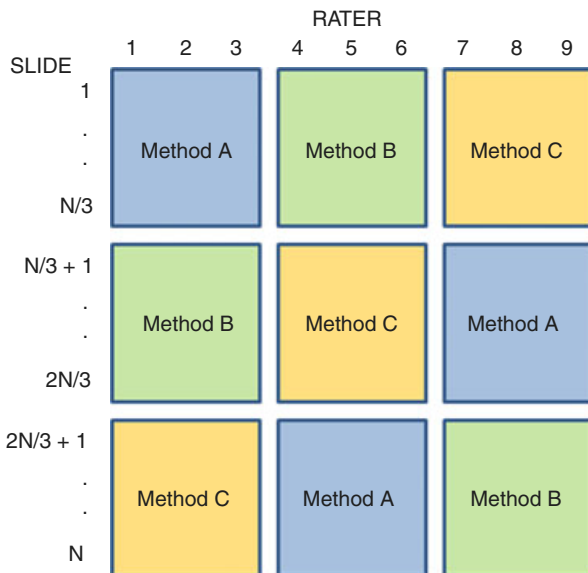


Figure 1. Study design. Each gynaecological pathologist was assigned a random number from 1 to 9. Observers 1–3 used method A (WHO classification) for the first third of the samples (block 1), method B (EIN classification) for the second third (block 2), and method C (EWG) for the last third (block 3). Methods for the first, second and third blocks of slides were B–C–A for observers 4–6 and C–A–B for observers 7–9.

Results

Table 1 shows the kappa values and 95% CIs for all major categories included in the WHO classification. One case was not evaluated by one of the observers (one missing evaluation). Full agreement (all three pathologists) and partial agreement (two pathologists)

Table 1. Kappa values, confidence intervals (95% CI) and full agreement between pathologists for every major category included in the WHO classification

	Kappa	(95% CI)	Full agreement (%)
Benign cycling endometrium	0.510	(0.399–0.621)	9.1
Simple hyperplasia without atypia	0.222	(0.115–0.329)	2.0
Complex hyperplasia without atypia	0.217	(0.098–0.336)	3.0
Simple hyperplasia with atypia	0.132	(–0.094 to 0.359)	0.0
Complex hyperplasia with atypia	0.299	(0.200–0.399)	4.6
Carcinoma	0.472	(0.363–0.581)	9.6

was obtained in 56 of 197 (28%) and 116 of 197 (59%) cases, respectively. Most disagreements (84/116; 72.4%) were mergers between adjoining categories (cycling endometrium versus simple hyperplasia without atypia; simple hyperplasia without atypia versus complex hyperplasia without atypia or simple hyperplasia with atypia; complex hyperplasia without atypia versus complex hyperplasia with atypia; simple hyperplasia with atypia versus complex hyperplasia with atypia; complex hyperplasia with atypia versus carcinoma or simple hyperplasia with atypia). However, 40% (44/110) of the discrepancies between two adjoining categories were disagreements between non-atypical hyperplasia and atypical hyperplasia, thus representing a diagnostic disagreement with a major clinical impact. Thirty-two of 116 (28%) discrepant diagnoses were jumps between distant categories (benign versus simple hyperplasia with atypia, complex hyperplasia without atypia, complex hyperplasia with atypia or carcinoma; simple hyperplasia without atypia versus complex hyperplasia with atypia or carcinoma; or complex hyperplasia without atypia versus carcinoma). Full disagreement between the three pathologists was observed in 25 of 197 (13%) samples.

Table 2 shows the kappa values and 95% CIs for all major categories included in the EIN classification.

Table 2. Kappa values, confidence intervals (95% CI) and full agreement between pathologists for every major category included in the endometrial Intraepithelial neoplasia (EIN) classification

	Kappa	(95% CI)	Full agreement (%)
Benign cycling endometrium	0.669	(0.5560–0.777)	11.2
Benign hyperplasia	0.349	(0.246–0.451)	9.7
Endometrial intraepithelial neoplasia	0.272	(0.181–0.364)	7.1
Carcinoma	0.522	(0.417–0.627)	11.2

Owing to broken slides, two cases were not evaluated by one of the observers (two missing evaluations). Full agreement and partial agreement was obtained in 77 of 196 (39%) and 110 of 196 (56%) cases, respectively. Most disagreements (103/110, 93.6%) occurred between adjoining categories (cycling endometrium versus benign hyperplasia; benign hyperplasia versus EIN; or EIN versus carcinoma), and only a small percentage (7/110; 6.4%) were skips between distant categories (cycling versus EIN or carcinoma; or benign hyperplasia versus carcinoma). However, 44 of 110 (40.0%) minor discrepancies between two adjoining categories were disagreements between benign hyperplasia and EIN, thus representing a diagnostic disagreement with a major clinical impact. Full disagreement between the three pathologists was observed in nine of 196 (5%) samples.

Table 3 shows the kappa values and 95% CIs for all major categories included in the EWG classification. Full and partial agreement was obtained in 117 of 198 (59%) and 74 of 198 (37%) cases, respectively. Most disagreements (71/74; 95.9%) occurred between adjoining categories (cycling endometrium versus hyperplasia; hyperplasia versus endometrioid neoplasia), and only a small percentage (3/74; 4.1%) were jumps between distant categories (benign versus endometrioid neoplasia). However, 47 of 74 (63.5%)

Table 3. Kappa values, confidence intervals (95% CI) and full agreement between pathologists for every major category included in the European Working Classification (EWG)

	Kappa	(95% CI)	Full agreement (%)
Benign cycling endometrium	0.626	(0.515–0.737)	10.2
Benign hyperplasia	0.398	(0.294–0.501)	13.7
Endometrioid neoplasia	0.621	(0.536–0.706)	35.5

minor discrepancies between two adjoining categories were disagreements between hyperplasia and endometrioid neoplasia, thus representing a diagnostic disagreement with a major clinical impact. Full disagreement between the three pathologists was observed in six of 198 (3%) samples.

In all three classifications, the kappa values for the extreme diagnoses (benign endometrium and carcinoma or endometrioid neoplasia) were higher than the kappa values obtained for the intermediate categories.

Table 4 shows the kappa values and 95% CIs of the three histological classifications included in the analysis. The results obtained with the diagnostic categories described for each classification, as well as the kappa values obtained after reducing the WHO classification categories into four, three and two categories, the EIN classification into three and two categories, and the EWG classification into two significant categories, are shown. In all three classifications, reduction of the number of diagnostic categories resulted in an increase in the kappa values. Only the EWG classification reduced to two categories reached a substantial level of agreement (0.621). Similarly, in all three classifications, reduction of the number of diagnostic categories resulted in an increase in full agreement between pathologists, which reached 70%, 69% and 72% for two categories of observations with the WHO, EIN and EWG classifications, respectively. Diagnostic trends for the nine pathologists participating in the study are shown in Figure 2.

To demonstrate in visual terms the challenges of these diagnoses, Figure 3A–E illustrates two cases with poor interobserver values, and Figure 3F shows a case with a good interobserver value.

Exclusion of the 39 lesions arising in polyps did not substantially alter the kappa values for each classification ($\kappa = 0.317$, 95% CI 0.255–0.380, for the

Table 4. Kappa values and 95% confidence intervals (95% CI) of the three histological classifications included in the analysis

	Kappa	(95% CI)	Full agreement (%)
WHO			
Six categories*	0.351	(0.293–0.410)	28
Four categories†	0.404	(0.338–0.470)	37
Three categories‡	0.391	(0.324–0.458)	37
Two categories§	0.591	(0.504–0.678)	70
EIN			
Four categories¶	0.434	(0.366–0.502)	39
Three categories**	0.528	(0.452–0.604)	58
Two categories***	0.589	(0.504–0.747)	69
EWG			
Three categories††	0.544	(0.465–0.622)	59
Two categories‡‡	0.621	(0.536–0.706)	72

WHO classification included the following categories:

*Six categories: 1, Benign; 2, simple hyperplasia without atypia; 3, complex hyperplasia without atypia; 4, simple hyperplasia with atypia; 5, complex hyperplasia with atypia; and 6, carcinoma.

†Four categories: 1, Benign; 2, hyperplasia without atypia (simple and complex); 3, atypical hyperplasia (simple and complex); and 4, carcinoma.

‡Three categories: 1, Benign; 2, hyperplasia without atypia; 3, atypical hyperplasia and carcinoma.

§Two categories: 1, Benign and hyperplasia without atypia; 2, atypical hyperplasia and carcinoma.

Endometrial intraepithelial neoplasia (EIN) classification included the following categories:

¶Four categories: 1, Cycling endometrium Benign; 2, benign hyperplasia; 3, E.I.N.; and 4, carcinoma.

**Three categories: 1, Benign; 2, benign hyperplasia; 3, E.I.N. and carcinoma.

***Two categories: 1, Benign and benign hyperplasia; 2, E.I.N. and carcinoma.

European working group (EWG) classification includes the following categories:

††Three categories: 1, Benign; 2, hyperplasia; 3, endometrioid neoplasia.

‡‡Two categories: 1, Benign and hyperplasia; 2, endometrioid neoplasia.

WHO classification with six categories; $\kappa = 0.464$, 95% CI 0.387–0.542, for the EIN classification with four categories; and $\kappa = 0.554$, 95% CI 0.469–0.640, for the EWG classification).

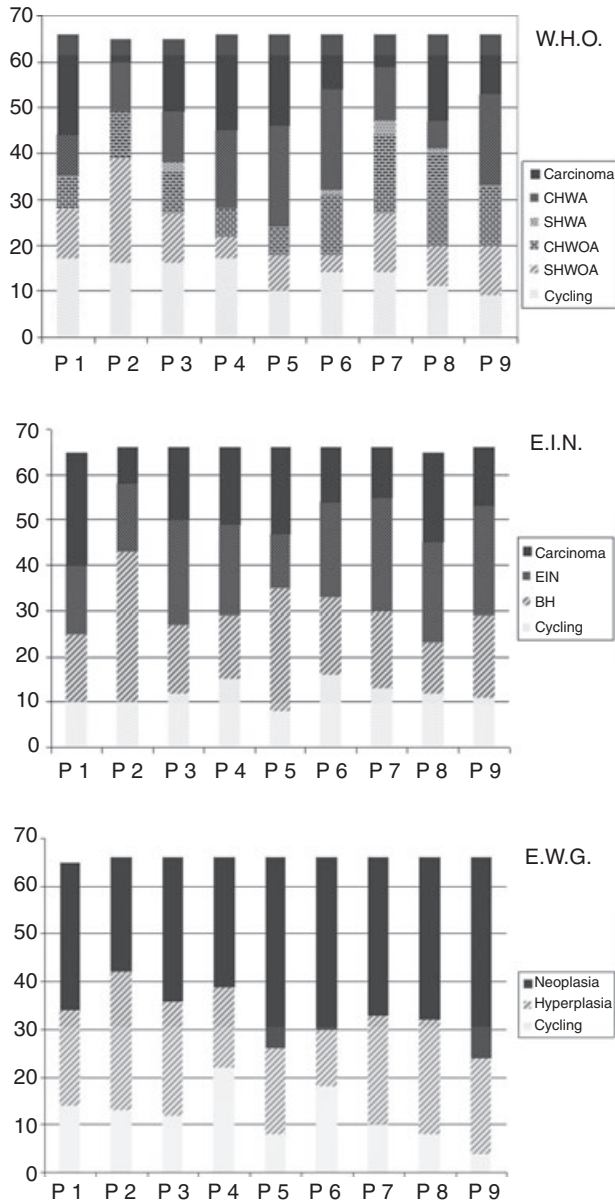


Figure 2. Diagnostic trends for the nine pathologists participating in the study. Top: WHO classification. Centre: endometrial intraepithelial neoplasia (EIN) classification. Bottom: European working group (EWG) classification. The graphs depict the allocation of diagnoses made by each of the nine panel pathologists (identification numbers are randomly assigned). CHWA, complex hyperplasia with atypia; SHWA, simple hyperplasia with atypia; CHWOA, complex hyperplasia without atypia; SHWOA, simple hyperplasia without atypia; EIN, endometrial intraepithelial neoplasia; BH, benign hyperplasia.

Discussion

This study confirms that all classifications of endometrial hyperplasia are associated with marked interob-

server variability, even among expert gynaecological pathologists. Indeed, only the EWG classification showed moderate agreement ($\kappa = 0.530$). It is relevant that complete agreement between all pathologists was observed in only slightly over one-quarter of the biopsies using the WHO classification, in one-third of the biopsies using the EIN classification, and in almost 60% of the biopsies using the EWG classification. As expected, in all three classifications, a reduction in the number of diagnostic categories resulted in increases in the kappa values and full agreement, but only the EWG classification reduced to two categories reached a substantial level of agreement (0.621).

The cases evaluated encompassed the full spectrum of endometrioid proliferative lesions. Our results are similar to those reported by Skov *et al.*,⁷ who looked at diagnostic agreement in the interpretation of 128 consecutive endometrial biopsies originally reported as endometrial hyperplasia among six gynaecological pathologists. Using a four-category classification (simple hyperplasia, complex hyperplasia, simple atypical hyperplasia, complex atypical hyperplasia), they reported mean kappa values of 0.20 and 0.25 in two rounds of reviews. In another study focusing exclusively on the reproducibility of atypical hyperplasia (AH) diagnoses, a three-member panel of pathologists reached full agreement with the original diagnosis of AH in only 15% of specimens.⁸ The pairwise agreement among panellists yielded kappa values ranging from 0.34 to 0.43.

Some other previous studies that have included a substantial number of cases originally diagnosed as benign or carcinoma have reported higher levels of interobserver agreement, because extreme categories are generally associated with better reproducibility. For example, Bergeron *et al.*⁵ assessed the diagnostic reproducibility in a set of 56 specimens, 55% of which were classified either as negative or carcinoma. Agreement among five European experts using an initial seven-category classification yielded a kappa value of 0.68, which improved to 0.76 when a three-category system was applied to the data.⁵ Kendall *et al.*⁶ performed a single-institution study in which five pathologists reviewed 100 specimens, 50 of which were originally diagnosed as negative or carcinoma. Reproducibility based on six categories (proliferative endometrium, four categories of endometrial hyperplasia, and carcinoma) yielded a kappa value of 0.69, with substantial intraobserver agreement.⁶

We assessed the potential to improve interobserver agreement by performing *post hoc* analyses in which WHO categories were merged. Agreement between

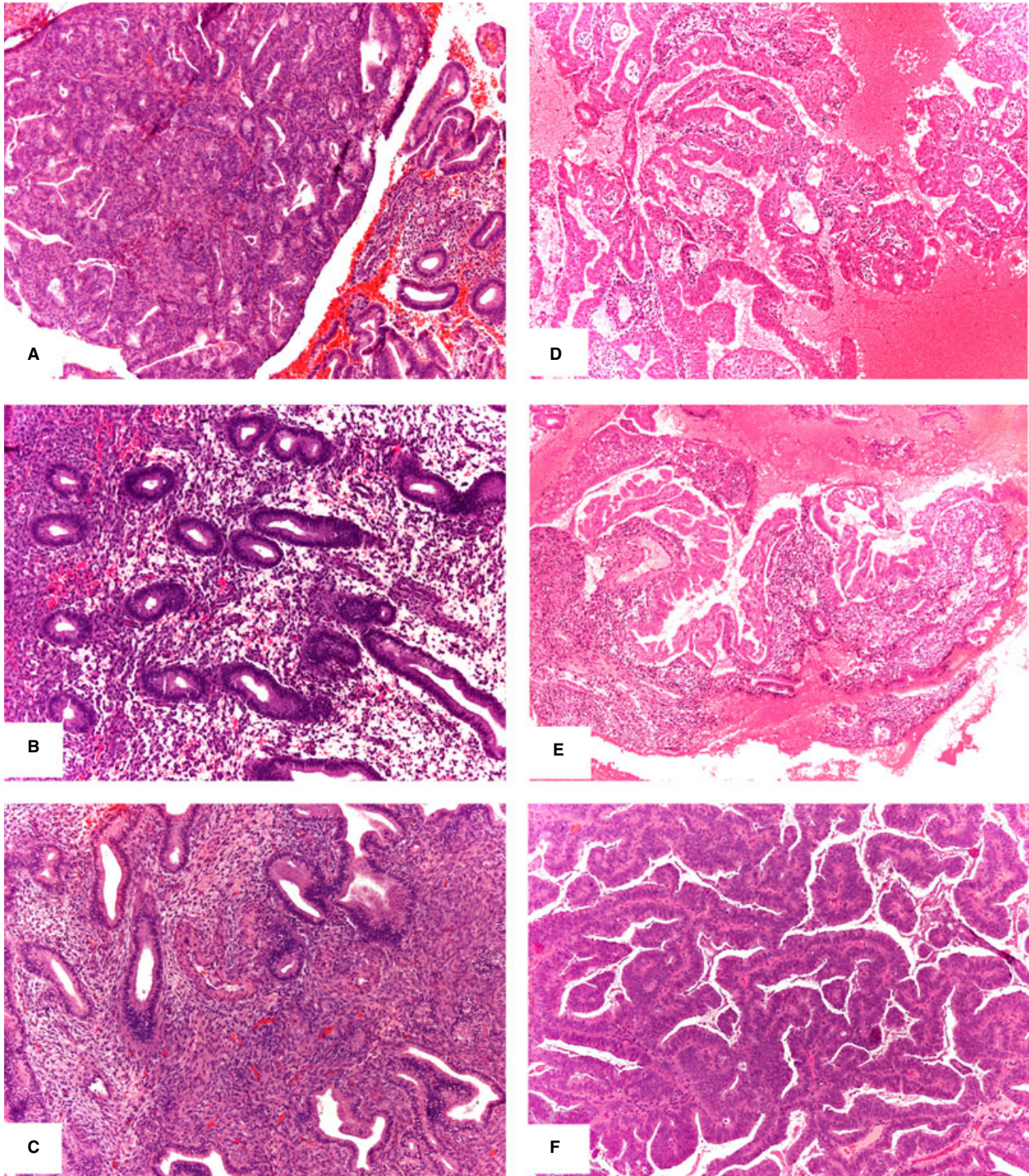


Figure 3. Histology of some cases with poor and good interobserver values. Case 44 (A–C) showed heterogeneous areas diagnosed as carcinoma (five observers), benign hyperplasia (one), complex hyperplasia (one), and benign (two). Diagnoses in case 112 (D,E) were carcinoma (one), benign hyperplasia (four), complex hyperplasia (three), and benign (two). All observers agreed in case 12 (F) on a diagnosis of carcinoma.

the original diagnoses and the panel was best for a two-tiered classification in which AH and carcinoma were separated from complex hyperplasia (CH) and less severe lesions. A dichotomous classification also improved the agreement between the panellists. Overall, the current reproducibility data suggest a revision of the WHO system, separating AH from less severe lesions. However, our focused morphological analysis has apparently failed to identify reliable criteria that greatly improve interobserver agreement in separating CH from AH or improve the assessment of progression risk. Therefore, other options for improving biopsy diagnosis of endometrial hyperplasia should be considered. Overall, the data on reproducibility suggest that one proposal for revising the WHO system is to dichotomize lesions by separating AH from less severe lesions. This has been incorporated in the upcoming 2014 WHO classification, where dichotomous categories of non-atypical and atypical hyperplasia (EIN) are used.

Proposals to simplify the WHO classification or to develop a new system have been previously published.^{5,9,17,18} Bergeron *et al.*⁵ suggested the term 'endometrioid neoplasia' for biopsies showing changes ranging from AH to grade 1 endometrioid carcinoma, because these lesions are inconsistently distinguished in biopsies, and are usually treated similarly with hysterectomy. Furthermore, even among women who seek uterine preservation or are at high operative risk because of comorbid factors, pathological distinction of AH from low-grade endometrioid carcinoma on biopsy may be of secondary importance, given that conservative management, usually hormonal, could be considered for either diagnosis, albeit with some risk. As shown in our study, thresholds for diagnosing low-grade endometrioid carcinoma on biopsy vary substantially. However, identifying high-grade endometrioid or non-endometrioid carcinomas and radiologically excluding myometrial invasion, metastases and synchronous ovarian carcinomas are always essential.

Instead of advocating revisions to the WHO classification, Mutter *et al.*⁹ have proposed a new classification, initially based on computer-assisted morphometric analysis and molecular testing. However, neither the WHO nor the EIN systems are easily or routinely applied strictly. Pathologists using WHO criteria are likely to use a reduced threshold for recognizing cytological atypia in the setting of glandular crowding. Our focused evaluation has demonstrated that assessing 'volume percentage stroma' (VPS), as required in the EIN system, is neither easy nor reproducible. Similarly to assessing atypia in the WHO

classification, assessment of VPS is limited by the pathologist's judgement of whether the biopsy contains a distinctive intact focus of glands where VPS should be assessed, and the estimate of the severity of crowding in such areas.

Our results, in combination with the data suggesting that carcinoma risk is considerably greater for AH than for other categories, suggest that a simplified dichotomous classification is desirable. One option would be to modify the WHO system for reporting biopsies (but not necessarily hysterectomy specimens) into two categories: (i) non-atypical endometrial hyperplasia; and (ii) AH/suspicious for grade 1 endometrioid carcinoma. However, focused efforts to define more specific, quantitative and reproducible histopathological criteria for defining a true cancer precursor would be helpful. In fact, a new system that borrows criteria from both the WHO and EIN classifications may be needed to achieve real diagnostic improvement.

Author contributions

J. Ordi, D. Hardisson and F. F. Nogales designed the study and selected the cases for review. C. Bergeron, W. G. McCluggage, H. Hollema, A. Felix, R. A. Sollow, E. Oliva, F. A. Tavassoli, I. Alvarado-Cabrero and M. Wells reviewed the slides. The manuscript was written by J. Ordi and F. F. Nogales. Each co-author corrected and amended the original text.

Acknowledgements

E. Quiñonez Urrego designed the histopathological illustrations. This work was presented in part at the Pathological Society of Great Britain and Ireland, Summer Meeting 2012, Sheffield, UK.

References

1. Null DB, Weiland CM, Camlibel AR. Postmenopausal bleeding – first steps in the workup. *J. Fam. Pract.* 2012; **61**: 597–604.
2. Sweet MG, Schmidt-Dalton TA, Weiss PM, Madsen KP. Evaluation and management of abnormal uterine bleeding in premenopausal women. *Am. Fam. Physician* 2012; **85**: 35–43.
3. Kurman RJ, Kaminski PF, Norris HJ. The behavior of endometrial hyperplasia. A long-term study of 'untreated' hyperplasia in 170 patients. *Cancer* 1985; **56**: 403–412.
4. Silverberg SG, Mutter GL, Kurman RJ, Kubik-Huch RA, Nogales FF, Tavassoli FA. Tumours of the uterine corpus. Epithelial tumours and related lesions. In Tavassoli FA, Devilee P eds. *World Health Organization classification of tumors. Pathology and genetics. Tumors of the breast and female genital organs. Tumors of the uterine corpus*. Lyon: IARC Press, 2003; 221–232.

5. 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**: 1102–1108.
6. Kendall BS, Ronnett BM, Isacson C *et al.* Reproducibility of the diagnosis of endometrial hyperplasia, atypical hyperplasia, and well-differentiated carcinoma. *Am. J. Surg. Pathol.* 1998; **22**: 1012–1019.
7. Skov BG, Broholm H, Engel U *et al.* Comparison of the reproducibility of the WHO classifications of 1975 and 1994 of endometrial hyperplasia. *Int. J. Gynecol. Pathol.* 1997; **16**: 33–37.
8. Zaino RJ, Kauderer J, Trimble CL *et al.* Reproducibility of the diagnosis of atypical endometrial hyperplasia: a Gynecologic Oncology Group study. *Cancer* 2006; **106**: 804–811.
9. Mutter GL, Baak JP, Crum CP, Richart RM, Ferenczy A, Faquin WC. Endometrial precancer diagnosis by histopathology, clonal analysis, and computerized morphometry. *J. Pathol.* 2000; **190**: 462–469.
10. Baak JP, Orbo A, van Diest PJ *et al.* Prospective multicenter evaluation of the morphometric D-score for prediction of the outcome of endometrial hyperplasias. *Am. J. Surg. Pathol.* 2001; **25**: 930–935.
11. Baak JP, Mutter GL. EIN and WHO94. *J. Clin. Pathol.* 2005; **58**: 1–6.
12. Mutter GL. Histopathology of genetically defined endometrial precancers. *Int. J. Gynecol. Pathol.* 2000; **19**: 301–309.
13. Mutter GL, Zaino RJ, Baak JP, Bentley RC, Robboy SJ. Benign endometrial hyperplasia sequence and endometrial intraepithelial neoplasia. *Int. J. Gynecol. Pathol.* 2007; **26**: 103–114.
14. Orbo A, Baak JP, Kleivan I *et al.* Computerised morphometrical analysis in endometrial hyperplasia for the prediction of cancer development. A long-term retrospective study from northern Norway. *J. Clin. Pathol.* 2000; **53**: 697–703.
15. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**: 159–174.
16. Lee J, Fung KP. Confidence interval of the kappa coefficient by bootstrap resampling. *Psychiatry Res.* 1993; **49**: 97–98.
17. Dietel M. The histological diagnosis of endometrial hyperplasia. Is there a need to simplify? *Virchows Arch.* 2001; **439**: 604–608.
18. Sherman ME, Ronnett BM, Ioffe OB *et al.* Reproducibility of biopsy diagnoses of endometrial hyperplasia: evidence supporting a simplified classification. *Int. J. Gynecol. Pathol.* 2008; **27**: 318–325.