

# Reproducibility of endometrial intraepithelial neoplasia diagnosis is good, but influenced by the diagnostic style of pathologists

Alp Usubutun<sup>1</sup>, George L Mutter<sup>2</sup>, Arzu Saglam<sup>1</sup>, Anil Dolgun<sup>3</sup>, Eylem Akar Ozkan<sup>4</sup>, Tan Ince<sup>5</sup>, Aytekin Akyol<sup>1</sup>, H Dilek Bulbul<sup>6</sup>, Zerrin Calay<sup>7</sup>, Funda Eren<sup>8</sup>, Derya Gumurdulu<sup>9</sup>, A Nihan Haberal<sup>4</sup>, Sennur Ilvan<sup>7</sup>, Seyda Karaveli<sup>10</sup>, Meral Koyuncuoglu<sup>11</sup>, Bahar Muezzinoglu<sup>12</sup>, Kamil H Muftuoglu<sup>6</sup>, Necmettin Ozdemir<sup>13</sup>, Ozlem Ozen<sup>4</sup>, Sema Baykara<sup>14</sup>, Elif Pestereli<sup>10</sup>, Emine Cagnur Ulukus<sup>11</sup> and Osman Zekioglu<sup>13</sup>

<sup>1</sup>Department of Pathology, Hacettepe University, Ankara, Turkey; <sup>2</sup>Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; <sup>3</sup>Department of Biostatistics, Hacettepe University, Ankara, Turkey; <sup>4</sup>Department of Pathology, Baskent University, Ankara, Turkey; <sup>5</sup>Department of Pathology, Braman Family Breast Cancer Institute and Interdisciplinary Stem Cell Institute, University of Miami Miller School of Medicine, Miami, FL, USA; <sup>6</sup>Etilik Zubeyde Hanim Women's Health Teaching and Research Hospital, Ankara, Turkey; <sup>7</sup>Department of Pathology, Istanbul University Cerrahpasa, Istanbul, Turkey; <sup>8</sup>Department of Pathology, Marmara University, Istanbul, Turkey; <sup>9</sup>Department of Pathology, Cukurova University, Adana, Turkey; <sup>10</sup>Department of Pathology, Akdeniz University, Antalya, Turkey; <sup>11</sup>Department of Pathology, Dokuz Eylül University, Izmir, Turkey; <sup>12</sup>Department of Pathology, Kocaeli University, Kocaeli, Turkey; <sup>13</sup>Department of Pathology, Ege University, Izmir, Turkey and <sup>14</sup>Department of Pathology, Uludag University, Bursa, Turkey

**Endometrial intraepithelial neoplasia (EIN) applies specific diagnostic criteria to designate a monoclonal endometrial preinvasive glandular proliferation known from previous studies to confer a 45-fold increased risk for endometrial cancer. In this international study we estimate accuracy and precision of EIN diagnosis among 20 reviewing pathologists in different practice environments, and with differing levels of experience and training. Sixty-two endometrial biopsies diagnosed as benign, EIN, or adenocarcinoma by consensus of two expert subspecialty pathologists were used as a reference comparison to assess diagnostic accuracy of 20 reviewing pathologists. Interobserver reproducibility among the 20 reviewers provided a measure of diagnostic precision. Before evaluating cases, observers were self-trained by reviewing published textbook and/or online EIN diagnostic guidelines. Demographics of the reviewing pathologists, and their impressions regarding implementation of EIN terminology were recorded. Seventy-nine percent of the 20 reviewing pathologists' diagnoses were exactly concordant with the expert consensus (accuracy). The interobserver weighted  $\kappa$  values of 3-class EIN scheme (benign, EIN, carcinoma) diagnoses between expert consensus and each of reviewing pathologists averaged 0.72 (reproducibility, or precision). Reviewing pathologists demonstrated one of three diagnostic styles, which varied in the repertoire of diagnoses commonly used, and their nonrandom response to potentially confounding diagnostic features such as endometrial polyp, altered differentiation, background hormonal effects, and technically poor preparations. EIN diagnostic strategies can be learned and implemented from standard teaching materials with a high degree of reproducibility, but is impacted by the personal diagnostic style of each pathologist in responding to potential diagnostic confounders.**

*Modern Pathology* (2012) 25, 877–884; doi:10.1038/modpathol.2011.220; published online 3 February 2012

**Keywords:** endometrial hyperplasia; endometrial intraepithelial neoplasia; reproducibility

Correspondence: Dr A Usubutun, MD, Hacettepe University, Medical Faculty, Department of Pathology, Sıhhiye, Ankara 06100, Turkey.  
E-mail: ausubutu@hacettepe.edu.tr

Received 12 October 2011; revised 25 November 2011; accepted 9 December 2011; published online 3 February 2012

Correct diagnosis of precancerous endometrial lesions enables treatment before progression to an invasive neoplasm. Correspondingly, exclusion of a premalignant lesion is important to avoid over treatment. Precancerous lesions in the current WHO schema are encompassed within a spectrum of four types of endometrial hyperplasias, diagnosis of which has a high intra- and interobserver variability<sup>1–5</sup> in addition to modest predictive value for carcinoma.<sup>1,2,6</sup> New diagnostic criteria, for fewer diagnostic categories, are required to resolve these issues.

Endometrial intraepithelial neoplasia (EIN) is a high-cancer-risk lesion first proposed by an international group of pathologists in 2000,<sup>7</sup> in response to newly emergent molecular, histomorphometric, and clinical outcome studies.<sup>7–9</sup> EIN fulfills all lines of evidence necessary to define a premalignant lesion, as established by a National Cancer Institute-sponsored consensus conference in 2006.<sup>10</sup> Molecular studies have shown that these lesions are clonally related to invasive carcinomas.<sup>11–13</sup> Correlation of molecular, histomorphometric, and clonal characteristics of endometrial lesions with clinical cancer outcomes has permitted *de novo* discovery of those diagnostic criteria, which characterize premalignant lesions, thus enabling precise definition of EIN diagnostic criteria.<sup>14,15</sup> Key changes include objective architectural thresholds (gland area exceeds that of stroma), interpretation of cytological changes relative to background normal glands in the same sample, and minimum lesion size (1 mm) within a single tissue fragment. EIN classification has better diagnostic reproducibility<sup>16</sup> and cancer predictive value than<sup>17</sup> WHO endometrial hyperplasias. Patients with EIN lesions have a 45-fold increased endometrial cancer risk.<sup>13</sup>

EIN has simple diagnostic criteria, outlined above, that can be used on a routine basis using only standard hematoxylin and eosin-stained tissue without the need for additional immunohistochemical or sophisticated molecular studies.<sup>8</sup> This characteristic makes it very appealing to the routine pathologist. Successful implementation of any new diagnostic schema, however, requires learning new criteria and consistently applying them in the intended manner. This is especially challenging with EIN, as the background tissue appearance is highly variable, and some lesion features such as size, method of assessing cytological change, and threshold gland densities are unfamiliar to many pathologists. Prior studies testing the reproducibility of this terminology have involved teams consisting only of gynecological pathologists, moreover some of these were carried out using the morphometric *D*-score rather than plain light microscopy.<sup>9,18,19</sup> In this study we test the reproducibility of the EIN diagnosis using only light microscopy. We further tried to emphasize its simplicity, with respect to the ease of learning and application of the diagnostic criteria, even to the inexperienced routine pathologist.

## Materials and methods

### Case Selection

Sixty-two endometrial biopsy and curettage materials performed for suspicion of endometrial hyperplasia between the years of 2001 and 2008, were retrieved from the archives of the Pathology Department of the Hacettepe University. We compiled a case mix of roughly equal numbers of atypical ( $n=25$ ), complex non-atypical endometrial hyperplasia ( $n=21$ ) and simple non-atypical ( $n=4$ ) supplemented with examples diagnosed previously as diagnostically difficult benign lesions suspected by the pathologist to represent some form of hyperplasia ( $n=8$ ) or carcinoma ( $n=4$ ). One representative slide from each of the 62 technically adequate endometrial biopsies was assessed by the reviewing study pathologists.

### EIN Diagnostic Categories

Each case was classified into one of the following diagnostic entities: specimen technically inadequate for diagnosis, benign non-EIN, EIN, or adenocarcinoma. EIN was diagnosed according to published criteria that include presence of all five diagnostic elements as listed here:<sup>15</sup>

- (1) Gland area greater than stromal area
- (2) Cytological demarcation of the lesion relative to background
- (3) Lesion size exceeds 1 mm in minimum diameter
- (4) Exclusion of mimics
- (5) Exclusion of cancer

### Reference EIN Diagnoses by Expert Panel ('Expert Pathologists')

All 62 cases were rediagnosed using published EIN criteria by two subspecialty trained gynecological pathologists involved in the initial development of EIN diagnostic criteria (expert 1 and expert 2). Each expert independently reviewed all cases blinded to history and prior diagnoses, and a consensus expert diagnosis ('reference') was attained by blinded agreement, or in cases of disagreement, by adjudication at a multiheaded microscope.

### Review by Practicing Pathologists ('Reviewing Pathologists')

Following review of published training materials, 20 practicing pathologists rediagnosed all cases blinded to the reference EIN diagnosis and all clinical features. These 20 European pathologists differed in the duration of practice experience and practice context. Seventeen were gynecological pathologists working in different regions of Turkey either in university or community hospitals with

varying gynecological material workloads. Three were general surgical pathologists without specialized training in gynecopathology. Before evaluating the cases all participants were advised to read two references ; 'Benign Endometrial Hyperplasia and EIN 'chapter in Robboy's Pathology of Female Reproductive Tract'<sup>20</sup> and view online instructional material at [www.endometrium.org](http://www.endometrium.org). After scoring the slides, each pathologist completed a questionnaire grading (on a 5-point scale) the learnability and usefulness of the EIN terminology in daily practice (Table 1).

### Statistical Analysis

A measure of diagnostic accuracy for each reviewing pathologist was generated by comparison with reference diagnoses. Reference diagnoses were the adjudicated consensus diagnoses of experts 1 and 2. Accuracy was expressed in two metrics: that percentage of cases in which the reference and reviewing pathologist diagnoses were concordant; and the weighted kappa statistic ( $\kappa_w$ ) between the reviewing pathologist and reference diagnoses. Interobserver agreement between the 20 reviewing pathologists, a measure of diagnostic reproducibility, was calculated using the kappa statistic ( $\kappa$ ) for multiple raters when there are more than two diagnostic outcomes.<sup>21</sup>

To evaluate the effect of current institutional affiliation, type of practice (speciality gynecological

pathologist vs general surgical pathologist) and the diagnostic system currently used (EIN vs WHO) on the % agreement between reviewer and reference; we used Mann–Whitney *U* and Kruskal–Wallis tests. We explored the effect of the continuous variable of years of pathology experience on the % agreement using the Spearman correlation analysis. Kappa analyses and the statistical tests were performed in STATA version 10.0. The statistical significance was set at  $P < 0.05$ .

Diagnostic trends were examined by hierarchical cluster analysis in a heat-map (color = diagnosis) matrix of reviewer (columns) by case (*X* and *Y* axis, respectively). On the basis of the hierarchical clustering of 20 reviewers, each was assigned to a diagnostic style group. Cluster analysis using a metric of percentage agreement and Ward's linkage was performed in Systat (v13, 2009, Systat Software, Chicago, IL, USA).

## Results

### Specimen Adequacy

All 62 specimens were considered technically adequate by the expert reviewers, who assigned diagnoses to each. Among the 20 reviewers reviewing the 62 cases, there were 1240 'diagnostic passes'. Of these, 99% (1229/1240) were considered diagnosable by the reviewers, who rendered a specific diagnosis.

**Table 1** Answers of 20 reviewing pathologists to the questionnaire

Questions	Categories	Frequency	Percent
Year of experience in gyn. pathology or if general pathologist, years spend in general pathology	< 3 years	1	5
	3–10 years	5	25
	> 10 years	14	70
Work place	University	16	80
	Public hospital etc.	4	20
Number of endometrial biopsies seen in one month	< 10	2	10
	10–20	1	5
	> 20	17	85
Which terminology do you use for endometrial hyperplasias?	EIN	4	20
	WHO	16	80
Did you read the related chapter in <i>Robboy's Pathology of the Female Reproductive Tract</i> ?	No	2	10
	Yes	18	90
Did you visit the <a href="http://endometrium.org">endometrium.org</a> web site?	No	2	10
	Yes	18	90
Do you think that EIN classification is easy to learn?	Absolutely disagree	0	0
	Disagree	0	0
	I have no idea	3	15
	Agree	10	50
	Absolutely agree	7	35
Do you think that EIN classification is easy to apply in daily practice?	Absolutely disagree	0	0
	Disagree	0	0
	I have no idea	6	30
	Agree	12	60
	Absolutely agree	2	10

**Expert Review and Reference Consensus Diagnoses**

The two expert reviewers agreed on the diagnostic classification as EIN *vs* other (benign non-EIN, or adenocarcinoma) and the consensus expert reference diagnoses, included 27 benign non-EIN, 26 EIN, and 9 adenocarcinoma cases.

**Agreement Between Expert Reference and Each of 20 Observers**

The percent agreement with regard to the reference diagnoses for the 20 individual pathologists are shown in Table 2. Overall, 79% of all reviewer diagnoses were exactly concordant with the reference expert diagnoses. As expected, unanimous agreement among all 20 reviewers was uncommon, occurring in 18 cases, (17 benign, 1 adenocarcinoma). The weighted  $\kappa$  values of each observer with the final consensus diagnosis are shown in Table 2, averaged 0.72, and varied between 0.45 and 0.84.

**Interobserver Reproducibility Among 20 Independent Observers**

The extent of interobserver diagnostic concordance between all pairwise permutations of each

of the 20 reviewers was calculated as a  $20 \times 20$  Spearman correlation matrix of 62 cases across all reviewers. Of all 190 possible pairwise interobserver comparisons of reviewer classification, the average Spearman correlation coefficient was 0.72, range 0.37–0.93. The overall  $\kappa$  (Interobserver reproducibility) among 20 independent observers was calculated as 0.58 for all diagnostic groups (benign *vs* EIN *vs* cancer), 0.64 for benign (*vs* EIN + CA), 0.47 for EIN (*vs* benign + cancer), and 0.64 for carcinoma (*vs* benign + EIN).

**Characteristics of 20 Independent Reviewers**

Table 2 gives descriptive statistics of the 20 reviewing pathologists. The 20 reviewers were trained at 13 different institutions, and practicing at 11 different hospitals at the time of the study. In response to the questionnaire (Table 1) only one observer read none of the recommended text, the other observers read at least one (two observers) or both of the references before evaluating the cases. On a 5-grade scale (1 defines ‘disagree’, while 5 defines ‘absolutely agree’) observers think that the EIN terminology is learnable (mean 4.16) and moderately easy to use in daily practice (mean 3.76).

**Table 2** Agreement between each of the 20 reviewing pathologists and consensus expert reference diagnoses (Percent concordant, weighted kappa statistics,  $\kappa_w$ )

Reviewer	Years of pathology practice	Training institution	Practice institution	Practice type	Uses EIN yes/no	Dx style group	Agreement of each reviewer with expert consensus reference		
							Percent concordant	The weighted kappa statistic ( $\kappa_w$ )	95% Bootstrap CI for the $\kappa_w$
Expert 1	28	CUCPS	BWH	GYN	Yes	Red	96.72	NA	NA
Expert 2	14	BWH	BWH	GYN	Yes	Yellow	90.32	NA	NA
Consensus expert diagnoses (reference)	NA	NA	NA	NA	NA	NA	100	NA	NA
Reviewer A	13	ANEAH	EZH	GYN	No	Red	88.71	0.8351	0.707–0.944
Reviewer B	13	HNEAH	AUFM	GYN	No	Yellow	87.10	0.8225	0.701–0.925
Reviewer C	16	IUCFM	IUCFM	GYN	No	Yellow	86.89	0.8229	0.707–0.937
Reviewer D	12	GUFM	EZH	GYN	No	Yellow	85.48	0.7932	0.653–0.907
Reviewer F	6	UUFM	BUFM	GEN	Yes	Red	85.48	0.7734	0.614–0.900
Reviewer E	15	HUFM	HUFM	GYN	Yes	Red	83.87	0.7685	0.621–0.888
Reviewer I	18	IUCFM	IUCFM	GYN	No	Yellow	83.33	0.7734	0.644–0.907
Reviewer G	17	UUFM	UUFM	GYN	Yes	Yellow	83.05	0.7813	0.657–0.898
Reviewer H	33	EUFM	AUFM	GYN	No	Yellow	82.46	0.7656	0.617–0.889
Reviewer J	8	DEUFM	DEUFM	GYN	No	Yellow	80.65	0.7255	0.594–0.871
Reviewer K	8	CUFM	CUFM	GYN	No	Yellow	80.65	0.7429	0.603–0.870
Reviewer L	7	IUFM	KUFM	GYN	No	Red	79.03	0.6797	0.539–0.818
Reviewer M	1	HUFM	HUFM	GEN	No	Red	79.03	0.7180	0.572–0.847
Reviewer N	4	BUFM	BUFM	GYN	No	Green	75.81	0.6809	0.518–0.825
Reviewer P	5	HUFM	HUFM	GEN	Yes	Yellow	75.81	0.6873	0.549–0.838
Reviewer Q	21	DEUFM	DEUFM	GYN	No	Yellow	75.81	0.6904	0.561–0.824
Reviewer O	3	AEAH	BUFM	GYN	No	Yellow	74.19	0.6630	0.481–0.797
Reviewer R	12	MUFM	MUFM	GYN	No	Green	72.58	0.6276	0.450–0.787
Reviewer S	13	EUFM	EUFM	GYN	No	Green	69.35	0.5109	0.351–0.716
Reviewer T	25	EUFM	EUFM	GYN	No	Green	66.13	0.4521	0.274–0.637

All kappa values were significant at the  $P < 0.001$  level. Experts 1 and 2 contributed reference diagnoses by adjudicated consensus. Pathologist demographics are shown, including years of practice, acronyms of the training institution, acronyms of the practice institution at time of study, practice type (speciality gynecological pathologist ‘GYN’, or general surgical pathologist ‘GEN’), current use of EIN in practice, and diagnostic style group (from Figure 1). NA: not applicable.

The extent of concordance between reviewer and expert reference diagnoses (Table 2) was not significantly different by reviewer years of experience ( $r=0.149$ ,  $P=0.529$ ), practice type ( $P=0.926$ ), training institution ( $P=0.082$ ), practice institution at the time of study ( $P=0.255$ ), or classification system used in practice ( $P=0.437$ ).

### Diagnostic Style Groups

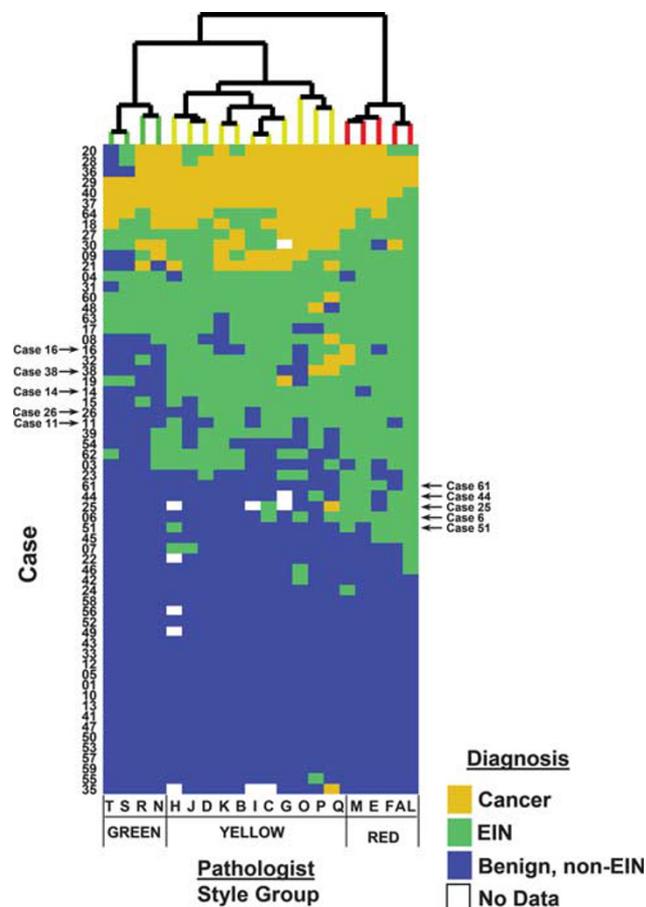
Figure 1 shows patterns of pathologist diagnosis based on percentage of matching interobserver diagnoses. Discordant cases (Figure 1, rows, marked by arrows), which polarized the reviewing pathologists (Columns) into different diagnostic clusters (diagnostic ‘style groups’) groups were identified (Figure 1, arrows).

Figure 1 is a detailed hierarchical dendrogram organizing each pathologist into one of three distinct diagnostic style groups designated here by the color of the branch termini (green, yellow, red). When the expert reference diagnoses are reintroduced in the same analysis by comparison, the reference diagnoses cosegregate with the red style. Diagnostic style group membership was not significantly associated with reviewer years of experience ( $P=0.435$ ), practice type (exact  $P=0.228$ ), training institution (exact  $P=0.236$ ), practice institution at the time of study (exact  $P=0.204$ ), or classification system used in practice (exact  $P=0.376$ ).

Unblinded secondary review was undertaken by one expert pathologist (expert 1) and the study head (reviewer E) for subsets of cases (Figure 1, arrows), which most distinguished the minority red and green diagnostic style groups (Figure 1). This included five cases preferentially diagnosed as EIN by the red group but benign by the others (Figure 1, arrows on the right), and five cases preferentially diagnosed as benign non-EIN by the green group, but as EIN by the others (Figure 1, arrows on the left). Unblinded observations from the secondary review of these cases are shown in Table 3.

### Discussion

This study shows good interobserver diagnostic reproducibility by a mixed group of pathologists applying published criteria for diagnosis of endometrial biopsies (cases with a previous diagnosis of hyperplasia) according to the EIN schema (benign, EIN, cancer). Benign, premalignant (EIN), and malignant (cancer) diagnoses represent an ascending severity of disease that initiate differing clinical interventions, underscoring the need to discriminate all three entities. A full-spectrum case mix permitted assessment of diagnostic reproducibility across two different decision thresholds of benign vs neoplastic (EIN or cancer) and non-malignant



**Figure 1** Diagnostic style groups of 20 independent reviewers. Hierarchical clustering of diagnoses rendered (tile color blue=benign, green=EIN, gold=cancer) in 62 cases (rows) by each of 20 reviewing pathologist (columns). Reviewer assignment to one of three major diagnostic style groups is shown at the top by a colored branching tree: green ( $n=4$ ), yellow ( $n=11$ ), and red ( $n=5$ ). Expert reference diagnoses cosegregate with the red cluster (not shown). Marked rows (arrows) are example case diagnostic discordances that distinguish the green (Cases 11, 26, 14, 38, 16) and red (Cases 51, 6, 25, 44, 61) style groups from the majority yellow group. Hierarchical dendrogram calculated using Wards Linkage, percentage matching distance metric.

(benign or EIN). Interobserver diagnostic reproducibility for both thresholds was very good, each with  $\kappa$ s of 0.64 across all 20 observers. This shows a balanced reproducibility between EIN and the range of lesser and worse entities with which it is likely to be confused.

Previous studies testing the reproducibility of the WHO hyperplasia classification have revealed poor to good interobserver agreement for the high-risk group (Table 4).<sup>1,2,5,22,23</sup> A peculiar feature is that, although the 1994 WHO hyperplasia classification contains four discrete hyperplasia entities (simple/complex, combined with atypical/non-atypical), few of these studies record and analyze four categories exactly as defined by the WHO. Atypical simple hyperplasia has such a low interobserver reproducibility ( $\kappa$  0.06–0.08)<sup>5</sup> and low frequency of

**Table 3** Unblinded review of selected cases that distinguish diagnostic styles

Case	Reference Dx	Red style Dx	Green style Dx	Unblinded comments (GM, AU)
51	EIN	EIN	Benign	EIN in EMP with morules Cytological changes subtle
6	EIN	EIN	Benign	EIN, 1 mm focus
25	EIN	EIN	Benign	EIN, focal lesion in a fragmented sample Breakdown present
44	Benign	EIN	Benign	Technically poor. Fragmented, shattered, thick
61	Benign	EIN	Benign	Thick sections obscure cytology Benign, anovulatory
11	EIN	EIN	Benign	EIN. Tubal differentiation Subtle cytological change
26	EIN	EIN	Benign	Localized EIN in anovulatory background Tubal differentiation
14	EIN	EIN	Benign	EIN. Extensive, loose in areas Tubal cytology, subtle change
38	EIN	EIN	Benign	EIN. Extensive. Tubal cytology
16	EIN	EIN	Benign	EIN in EMP

EIN, endometrial intraepithelial neoplasia; EMP, endometrial polyp.

**Table 4** Summary of the previous studies analyzing the reproducibility of the endometrial hyperplasia diagnosis

Study by	# Cases	# Reviewers	Diagnostic groups	Kappa value for interobserver agreement	
				Overall value across all diagnostic groups	Value for the high risk group yes/no
Zaino <i>et al</i> <sup>22, a</sup>	302	3	Normal, EH, AEH, cancer	0.40	0.28 (AEH)
Skov <i>et al</i> <sup>5, a</sup>	128	6	SEH, CEH, CAEH, SAEH	0.20–0.25	0.42–0.59 (ACEH) 0.06–0.08 (ASEH)
Skov <i>et al</i> <sup>5, a</sup>	128	6	ACEH vs other	0.42–0.49	0.42–0.49 (ACEH)
Kendall <i>et al</i> <sup>2, b</sup>	100	5	Normal, EH, AEH, cancer	0.69	0.47 (AEH)
Bergeron <i>et al</i> <sup>1, b</sup>	56	5	Normal, EH, AEH, cancer	0.52	
Bergeron <i>et al</i> <sup>1, b</sup>	56	5	Normal, EH, NEO	0.63	
Hecht <i>et al</i> <sup>16</sup>	97	3	Benign, EIN	0.54–0.62	0.54–0.62 (EIN)
Present study <sup>b</sup>	62	20	Benign, EIN, cancer	0.58	0.47 (EIN)

AEH, atypical (simple or complex) endometrial hyperplasia; EH, non-atypical (simple or complex) endometrial hyperplasia; CEH, non-atypical complex endometrial hyperplasia; SEH, non-atypical simple hyperplasia; ACEH, atypical complex endometrial hyperplasia, ASEH, atypical simple endometrial hyperplasia; NEO, combined AEH and cancer, EIN, endometrial intraepithelial neoplasia.

<sup>a</sup>Classification of cases with a diagnosis of hyperplasia.

<sup>b</sup>Classification of cases representing the entire spectrum of proliferative lesions of the endometrium.

occurrence that some groups doubt whether it truly exists<sup>5</sup> (Skov), whereas others beg the question by combining it with complex atypical hyperplasias during analysis. The result is that few of the studies use the 4-class WHO scheme as published and most published results represent contractions thereof. In its most extreme form, aggregation of complex and simple atypical hyperplasias with well differentiated carcinoma under the rubric of ‘endometrial neoplasia’ has been proposed.<sup>1</sup> In general, these aggregation strategies do improve interobserver reproducibility, there being fewer choices to consider. Thus the low overall interobserver reproducibility for the full 4-class WHO scheme ( $\kappa$  0.2–0.25),<sup>5</sup> improves substantially upon combining simple and complex atypical hyperplasias into a single group ( $\kappa$  0.40–0.69).<sup>1,2,22</sup> In these cases, what is actually

being reported upon is something different than the published WHO 4-class system. A further factor is that the overall reproducibility likelihood is affected by the range of non-hyperplastic tissues included in the series, especially whether carcinomas and normal tissues are included or excluded during case selection.

The EIN lesion is a singular entity representing a high-risk premalignant lesion, which from its inception has been documented by a published body of molecular, histopathological, and clinical outcome evidence. Similar to the aggregated WHO hyperplasia categories, it has the reproducibility advantage of fewer entities, but differs in achieving this by design and precise primary definition rather than secondary bundling of categories during analysis. Most pathologists now recognize that

the four histopathological WHO hyperplasias do not represent four different clinico-biological groups.<sup>1,2,7,22</sup> The evidence base for EIN includes new clinical outcome predictive diagnostic features such as lesion size, quantitative gland density, and methods of interpreting cytology that provide pathologists with improved diagnostic criteria relative to those within the 1994-vintage WHO scheme. Implementation of a tripartite (benign-EIN-cancer) scheme is most responsibly accomplished by using expanded and explicit criteria such as those published for EIN.

We carried forward the diagnostic patterns of each pathologist through complete data analysis, allowing us to discern how individual cases divided pathologists into diagnostic clusters or 'styles'. This is shown intuitively as a graphic matrix of cases by pathologists organized into clusters based on diagnostic proximity (Figure 1). Pathologists fell into three different style groups, which can simply be described as follows: (1) a majority group of 11 pathologists who used a balanced spectrum of benign, EIN, and carcinoma diagnoses (Figure 1, Yellow); (2) four pathologists who favored benign diagnoses and infrequently diagnosed EIN or carcinoma (Figure 1, green); and (3) five pathologists who favored EIN diagnosis more than their peers (Figure 1, red). Pathologists are well aware of this phenomenon of diagnostic style, using terms such as 'splitters vs lumpers,' or 'benign vs malignant' to classify the diagnostic personality of colleagues. These data demonstrate its expression in practice and suggest that diagnostic behavior of the individual pathologist is subject to limited modification by simple elaboration of diagnostic criteria or retraining. There are many possible personal experiences that might contribute to individual diagnostic style, but in this limited study we were unable to show a correlation with place of training, practice environment, years of practice, or practice type.

Detailed examination of those individual cases capable of polarizing pathologists into style groups provides clues to the basis of predictable, nonrandom diagnostic disagreement. The red style group had a tendency to diagnose reference EIN lesions not appreciated by the other groups, including those with confounding factors of altered differentiation (case 51), small size of localized EIN focus (case 6), or background stromal breakdown (case 25). Correspondingly, the red style group overdiagnosed as EIN some reference benign cases, which were technically inferior preparations (case 44, 61) or had confusing global estrogen-induced changes (case 61). The green style group tended to diagnose EIN reference cases as benign when tubal differentiation was geographically present in the EIN (cases 11, 14, 26, 38), when the EIN was focally distributed within only one of many fragments (case 26) or was present within a polyp (case 16).

The cases with discordant results were usually due to reasons previously described in the literature.<sup>8,16</sup>

EIN lesions may arise completely or partially within endometrial polyps and the diagnosis of EIN within a polyp can be problematic<sup>24</sup> because polyps themselves have variable gland density and somewhat altered gland cytology.<sup>8,16</sup> It has been stressed before that when making a diagnosis of EIN within a polyp the suspicious focus should be compared with the polyp background and not that of the uninvolved endometrium, so as not to mistake the somewhat altered gland cytology that can be seen within polyps for EIN.<sup>25</sup> The minimum size for a lesion to be considered as EIN was determined to be 1 mm in maximum dimension within a single tissue fragment.<sup>18,26,27</sup> Small foci can be overlooked, also accounting for discordance in our study. Technical artifacts, due to poor fixation and processing, in some cases made it difficult to evaluate both large scale (architecture) and small scale (cytology) features of EIN lesions. This is a problem not specific to the endometrium. Lastly, the altered cytology that characterizes EIN must be interpreted within the context of both cytoplasmic and nuclear findings. Several cases with prominent cytoplasmic changes were construed as benign metaplasias rather than EIN. The clue to an EIN diagnosis in these examples is the clustered and discrete nature of the altered gland focus, in contrast to randomly distributed hormonally induced metaplasias, or reactive cytological change in association with stromal breakdown.<sup>25</sup>

An important finding of our study is that EIN morphological criteria can be easily learnt and that learning ability is not affected by years of experience. One of the reviewers in the study, who was not a gynecological pathologist and was unfamiliar with the terminology and diagnostic criteria, had a good ( $\kappa$  0.72) diagnostic agreement with the reference diagnoses. Furthermore, survey results showed that the EIN morphological criteria are easy to learn (graded as 4.16, with 5.0 being 'agree completely'). These results underscore that self-learning of morphological criteria of EIN is possible and easy. However, reviewing pathologists thought that application of this knowledge to practice is not as easy as learning the criteria. Observers graded ease of application of the diagnostic criteria to daily practice as 3.76 on a 5-grade scale. The pathologists who were using the EIN terminology and diagnostic criteria on their daily practice had similar results and good agreement with the consensus diagnosis (0.69–0.78). We must remember that all of the WHO studies mentioned above use criteria, which every pathologist is familiar with. We may assume that the diagnostic reproducibility of the pathologists will increase as EIN criteria are used in daily practice, as results of our study suggest.

In summary the reproducibility of the EIN diagnoses was high in our study, especially considering the fact that the group consisted of a variety of pathologists. Our group included a mix of speciality gynecological and general pathologists, and we

were still able to reach  $\kappa$  values for interobserver agreement of moderate to substantial degree. The high interobserver reproducibility under these circumstances underlines the fact that this terminology is relatively simple, can easily be learned and applied by the inexperienced pathologist and hence has a potential for widespread appeal.

## Acknowledgement

We extend our gratitude to Havva Özşeker for assistance with the cases.

## Disclosure/conflict of interest

The authors declare no conflict of interest.

## References

- 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.
- 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.
- Dietel M. The histological diagnosis of endometrial hyperplasia - Is there a need to simplify? *Virchows Arch* 2001;439:604–608.
- Zaino RJ. Endometrial hyperplasia: is it time for a quantum leap to a new classification? *Int J Gynecol Pathol* 2000;19:314–321.
- 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.
- 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.
- Mutter GL. Endometrial intraepithelial neoplasia (EIN): Will it bring order to chaos? *Gynecol Oncol* 2000;76:287–290.
- Mutter GL, Zaino RJ, Baak JPA, *et al*. Benign endometrial hyperplasia sequence and endometrial intraepithelial neoplasia. *Int J Gynecol Pathol* 2007;26:103–114.
- Mutter GL, Baak JPA, Crum CP, *et al*. Endometrial precancer diagnosis by histopathology, clonal analysis, and computerized morphometry. *J Pathol* 2000;190:462–469.
- Berman JJ, Albores-Saavedra J, Bostwick D, *et al*. Precancer: A conceptual working definition results of a consensus conference. *Cancer Detect Prev* 2006;30:387–394.
- Mutter GL, Boynton KA, Faquin WC, *et al*. Allelotype mapping of unstable microsatellites establishes direct lineage continuity between endometrial precancers and cancer. *Cancer Res* 1996;56:4483–4486.
- Mutter GL, Chaponot ML, Fletcher JA. A polymerase chain-reaction assay for nonrandom x-chromosome inactivation identifies monoclonal endometrial cancers and precancers. *Am J Pathol* 1995;146:501–508.
- Hecht JL, Mutter GL. Molecular and pathologic aspects of endometrial carcinogenesis. *J Clin Oncol* 2006;24:4783–4791.
- Mutter GL. Histopathology of genetically defined endometrial precancers. *Int J Gynecol Pathol* 2000;19:301–309.
- Mutter GL. Diagnosis of premalignant endometrial disease. *J Clin Pathol* 2002;55:326–331.
- Hecht JL, Ince TA, Baak JPA, *et al*. Prediction of endometrial carcinoma by subjective endometrial intraepithelial neoplasia diagnosis. *Modern Pathol* 2005;18:324–330.
- Baak JP, Mutter GL, Robboy S, *et al*. The molecular genetics and morphometry-based Endometrial Intraepithelial Neoplasia classification system predicts disease progression in Endometrial hyperplasia more accurately than the 1994 World Health Organization classification system. *Cancer* 2005;103:2304–2312.
- Orbo A, Baak JPA, 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.
- Baak JPA, 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.
- Mutter GL, Zaino RJ, Baak JPA, *et al*. Benign endometrial hyperplasia and EIN, In: Robboy SJ, Mutter GL, Prat J, Bentley RC, Russel P, Anderson MC (eds). *Robboy's Pathology of Female Reproductive Tract*. 2nd edn. Elsevier: China, 2009, pp 367–397.
- Fleiss JL, Levin B, Paik MC, (eds). *Statistical Methods for Rates and Proportions*, 3rd edn. John Wiley & Sons: New York, 2003, pp 614–617.
- 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.
- Landis JR, Koch GG. Measurement of Observer Agreement for Categorical Data. *Biometrics* 1977;33:159–174.
- Mutter GL, Duska L, Crum CP. Endometrial intraepithelial neoplasia, In: Crum CP, Lee K (eds). *Diagnostic Gynecologic and Obstetric Pathology*. 1st edn. Saunders: China, 2006, pp 493–518.
- Carlson JW, Mutter GL. Endometrial intraepithelial neoplasia is associated with polyps and frequently has metaplastic change. *Histopathology* 2008;53:325–332.
- 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;174:1518–1521.
- Baak JPA, Nauta JJP, Wissebrekelmans ECM, *et al*. Architectural and nuclear morphometrical features together are more important prognosticators in endometrial hyperplasias than nuclear morphometrical features alone. *J Pathol* 1988;154:335–341.