Published in final edited form as:

Am J Surg Pathol. 2008 May ; 32(5): 691–698. doi:10.1097/PAS.0b013e318159a2a0.

Diagnosing Endometrial Hyperplasia: Why is it so difficult to agree?

KH Allison¹, SD Reed^{2,3,4}, LF Voigt^{3,4}, CD Jordan¹, KM Newton^{3,5}, and RL Garcia¹

¹Department of Pathology, University of Washington Medical Center

²Department of Obstetrics and Gynecology, University of Washington Medical Center

³Department of Epidemiology, University of Washington Medical Center

⁴Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA,

⁵Group Health Center for Health Studies, Seattle, WA

Abstract

Current WHO classification of endometrial hyperplasia is problematic because of poor diagnostic reproducibility. We sought to determine factors that cause diagnostic disagreement in a review of 2,601 endometrial specimens. Blinded random specimens of normal endometrium, hyperplasias and carcinoma were reviewed by two pathologists, with review by a third pathologist in cases with disagreement. All cases of endometrial hyperplasia or carcinoma were scored for degree of glandular crowding, architectural complexity and cytologic atypia. Sample adequacy, hyperplasia volume, presence of metaplasia or endometrial polyp were also scored. The overall kappa for agreement was 0.71, with a lower kappa of 0.36 when cases called “no hyperplasia” were excluded. The percent specific agreement was 90.3% for “no hyperplasia,” 31.1% for simple hyperplasia, 51.1% for complex hyperplasia, 49.8% for atypical hyperplasia and 57.5% for adenocarcinoma. Cases categorized as “low volume hyperplasia” had more diagnostic disagreement than “high volume,” (62% versus 39%, $p = 0.003$). Similarly, cases called “scant” had more diagnostic disagreement than “not scant” (65% versus 57%, $p = 0.013$). The histologic feature associated with the most diagnostic disagreement was cytologic atypia ($p < 0.0001$). Architectural crowding, architectural complexity or the presence of a polyp were all associated with diagnostic disagreement ($p < 0.0001$). High diagnostic disagreement in endometrial hyperplasia is related to both sample adequacy and interpretation of histologic features present. While obtaining additional tissue may increase diagnostic reproducibility, differences in interpretation of key histologic features like cytologic atypia remain major factors contributing to diagnostic disagreement.

Keywords

endometrial hyperplasia; diagnostic disagreement; endometrium; scant endometrial tissue; endometrial polyps; endometrial metaplasia

Introduction

Various classification schemes and terminology have been applied to the endometrium, all of which have relatively poor diagnostic reproducibility.^{2, 3, 5, 6, 8, 11, 20, 23} The currently accepted WHO terminology separates endometrial proliferations into simple or complex hyperplasia on the basis of architectural features and typical or atypical on the basis of cytologic features as originally defined by Kurman et al. in 1985.¹¹ This terminology was adopted by the WHO because of a reported increased risk of progression of lesions classified as complex

hyperplasia with atypia to carcinoma in contrast to lesions diagnosed as hyperplasia without atypia (23% of cases of atypical hyperplasias progressed while only 2% of those without atypia progressed, within a mean follow-up of 13.4 years).¹¹ Other studies have also reported the highest risks of progression to carcinoma in the atypical hyperplasia group, as well as the highest risk of persistence despite hormonal therapy.⁹ A recent prospective study also found that 43% of women with a diagnosis of atypical hyperplasia had concurrent carcinoma in hysterectomies performed within 12 weeks of diagnosis with no intervening therapy.²² These results suggest that a diagnosis of “hyperplasia with atypia” is a reliable predictor of women at high risk for subsequent or concurrent endometrial carcinoma. Currently, in the United States, a diagnosis of atypical hyperplasia usually leads to a recommendation of hysterectomy, rather than a trial of hormonal therapy. However, how reliably and how reproducibly can pathologists make this diagnosis?

Previous studies evaluating the diagnostic reproducibility of the 1994 WHO system (the categories of which remained unchanged in the current 2003 WHO system), reported kappa values for overall inter-observer agreement in diagnosing endometrial hyperplasia ranging from 0.2-0.7.^{1, 10, 19, 24} The diagnosis of atypical hyperplasia was found to be the least reproducible category, with kappa values ranging from 0.28-0.65 or percent agreement as low as 25%. Various reasons for the poor reproducibility of these categories have been proposed including variably applied criteria for the diagnosis of atypia, limited samples, and complicating features such as metaplasia or polyps. But importantly, specific factors involved in diagnostic disagreement have not been systematically evaluated. In addition, some of the previous studies have suffered from inadequate documentation of methodology and low numbers of cases reviewed.¹¹

As part of an ongoing cohort study of 1,799 women diagnosed by community pathologists with possible complex endometrial hyperplasia with or without atypia (identified through automated pathology records), we reviewed the index biopsy and subsequent endometrial samples (including normal endometrium, simple, complex, atypical hyperplasia and carcinoma) for 2,601 specimens. Diagnostic agreement was evaluated between review panel pathologists. Each case was scored for a variety of factors related to specimen quality, quantity and diagnostic criteria to elucidate features associated with diagnostic disagreement.

Materials and Methods

As part of the Endometrial Hyperplasia Outcomes (ECO) cohort study, 2,601 endometrial specimens underwent pathology review. Women aged 18-88 years, with a possible diagnosis of complex endometrial hyperplasia with or without atypia, were identified through records from January 1, 1985 through April 1, 2005 from automated pathology databases at Group Health (GH) in Washington State. To maximize the number of hyperplasia specimens and mimic how patients were most likely to be clinically classified, diagnoses with indefinite wording such as “cannot rule out” or “at least” were included in the pathology review and categorized as the higher grade diagnosis, with the exception of specimens with a diagnosis of “atypia cannot rule out low grade carcinoma,” which were included as possible atypical hyperplasias. Women initially diagnosed with endometrial carcinoma were excluded. Both index diagnostic specimens (1799 specimens) and all follow-up specimens (702 specimens) were included in the randomized, blinded pathology review. A total of 2,418 specimens were biopsies or curettages and 183 were hysterectomies.

Each case was independently reviewed by two pathologists, with additional independent review by a third pathologist in cases with diagnostic disagreement. Cases were assigned a final diagnosis based on majority diagnosis. If there was no majority diagnosis, the senior pathologist (RG) reviewed the three diagnoses and selected the “middle category” diagnosis.

Final diagnosis was classified as simple hyperplasia, complex hyperplasia, atypical hyperplasia, carcinoma, or “no hyperplasia” (specified as proliferative, normal secretory, atrophic/inactive, or shedding/menstrual). Carcinomas were graded according to the FIGO system.¹⁸ All cases diagnosed using WHO criteria as endometrial hyperplasia or endometrial carcinoma were scored for degree of glandular crowding, architectural complexity and cytologic atypia. In addition, sample adequacy, volume of hyperplastic tissue, and presence of metaplasia or endometrial polyp were noted. Appendix 1 summarizes the criteria used for this scoring system.

All three pathologists are academic pathologists. The two primary reviewers have a subspecialty focus in gynecologic pathology (KA and RG). RG and CJ have each had over ten years experience in practice and KA was a fellow in gynecologic and breast pathology at the study onset and became faculty during the study. Pathologists reviewed an initial pilot series of cases together to establish definitions for the features scored (Appendix 1) and pathologists were instructed to use the current WHO criteria in establishing a diagnosis; thereafter cases were reviewed independently. Initial pilot studies on 38 specimens for reviewer RG resulted in no intra-observer variability. Pathologists were blind to community-based diagnosis and all clinical information with the exception of patient age. Index and follow-up biopsies were randomly mixed for review.

Information regarding demographics (age and race), reproductive, medical and family history, and physical characteristics, including height and weight at the time of the index biopsy, were collected from the GH medical record; a single document containing all records from outpatient visits, test reports, and records of hospitalizations and consultations.

The kappa statistic was used to measure inter-reviewer agreement.¹² Kappa values were computed using STATA 9.2 (StataCorp, College Station, Texas). Percent agreement was computed as the number of specimens where the two reviewers agreed exactly divided by the total number of eligible specimens. Overall weighted kappa and weighted agreement were computed for by ordering the diagnoses as follows: no hyperplasia, simple hyperplasia, complex hyperplasia, hyperplasia with atypia and carcinoma. The following standard weights in STATA 9.2 were used: 1.0 for perfect agreement, 0.75 for adjacent categories, 0.5 for diagnosing two categories away from each other, 0.25 for those three categories away, and 0 for all others. Kappa values for individual diagnoses were computed by collapsing the data into two-by-two tables (agreed on a diagnosis, either reviewer disagreed on this diagnosis, both agreed the specimen was some other diagnosis) as described by Fleiss.⁴ Percent specific agreement was computed for individual diagnoses instead of percent agreement overall to avoid giving inappropriate weight to the large category in each table where both reviewers agreed the specimen was “some other diagnosis.”⁴ P values for the difference between proportions were computed using the PEPI program “Differ” (Abramson JH, Gahlinger, PM. Computer Programs for Epidemiologic Analyses: PEPI v. 4.0 J.H. Sagebrushpress, 2003; <http://www.sagebrushpress.com/pepibook.html>). In the data analysis for tables 4, 5 and 6, cases where both reviewers diagnosed “no hyperplasia” or either reviewer diagnosed “cannot rule out hyperplasia,” “inadequate sample,” or “non-diagnostic” were excluded because they were not scored for volume of hyperplasia or the presence of specific diagnostic criteria.

Results

A total of 1,799 women were included in the study cohort population. Clinical characteristics of the cohort are summarized in Table 1. The most common age range was 45-54 years. Most of the women were white and it was most common to have a body mass index of 30 kg/m² or greater.

The two primary pathologists reviewed a total of 2,601 specimens with 577 (22.2%) resulting in diagnostic disagreement which went to review by the third pathologist. Final panel diagnosis resulted in 1,829 “no hyperplasia or simple hyperplasia” diagnoses, 396 “complex hyperplasia”, 288 “atypical hyperplasia”, 54 “adenocarcinoma” and 34 “other” (inadequate, non-diagnostic or cannot rule out hyperplasia). (Table 2)

Overall diagnostic agreement between the two main panel pathologists was 91.1% (weighted percent agreement), with a weighted kappa of 0.71. (Table 3) Disagreements resulting in a third review were most frequent when one pathologist diagnosed complex hyperplasia and the other diagnosed complex hyperplasia with atypia (29.2% of cases with disagreement). Most disagreements were within one diagnostic category of each other, with a small percent (9.3%) of disagreements significant up or downgrades from “no hyperplasia” or simple hyperplasia to atypical hyperplasia or carcinoma or vice versa.

Diagnostic trends between the two primary pathologists are shown in Figure 1A. Pathologist A was more likely to diagnose atypical hyperplasia or carcinoma while pathologist B was more likely to diagnose complex, simple or no hyperplasia. But overall diagnostic trends were similar. Diagnostic trends for cases sent to pathologist C are shown in Figure 1B. Reviewer C agreed with 43.9% of pathologist B’s and 27.7 % of pathologist A’s diagnoses. Reviewer C also had a similar frequency of diagnosing carcinoma as reviewer B. However, reviewer C diagnosed atypical hyperplasia more often than reviewer B.

Diagnostic agreement by the panel for each WHO category is shown in Table 3. Agreement was highest for “no hyperplasia” and lowest for simple hyperplasia. Agreement for complex and atypical hyperplasia was fair (kappa = 0.21 and 0.35), and agreement for adenocarcinoma was moderate (kappa = 0.55). There were no statistically significant differences in agreement for index verses follow-up specimens.

Association of the quantity of diagnostic material present in each case with diagnostic disagreement is presented in Table 4. Cases with specimen adequacy scored by either pathologist as “scant” had more diagnostic disagreement than those scored “moderate” or “abundant” ($p = 0.013$). In addition, a low volume of hyperplastic tissue present was significantly associated with diagnostic disagreement with ($p < 0.0001$). Cases with a high volume of hyperplasia had the least amount of diagnostic disagreement (38.8%). Non-hysterectomy and hysterectomy specimens had similar diagnostic disagreement ($p = 0.51$).

Cases with diagnostic disagreement were more likely to disagree on all three histologic features evaluated than cases with diagnostic agreement. (Table 5) Almost half (47.3%) of the cases with diagnostic disagreement had disagreement on the degree of cytologic atypia, compared to 16.1% of the cases with diagnostic agreement ($p < 0.0001$). Cases with disagreement about the degree of glandular crowding included 33.5% of cases with diagnostic disagreement, compared to 21.1% of the cases with diagnostic agreement ($p < 0.0001$). Cases with disagreement about the degree of architectural complexity included 31.3% of cases with diagnostic disagreement, compared to 22.1% of the cases with diagnostic agreement, ($p = 0.005$). See Figure 2 for the histology of examples of cases with diagnostic disagreement.

“Complicating factors” such as the presence of a polyp or abundant metaplasia were also investigated for their effect on agreement. There was greater diagnostic disagreement when features of a polyp were noted by one of the pathologists. Of 230 cases where a polyp was noted to be present, 164 (71.3%) also had diagnostic disagreement compared with 521 of 931 cases (56%) where no polyp was noted ($p < 0.0001$). However there was only a suggestion of increased diagnostic disagreement when metaplastic changes were noted by one of the

pathologists (66.2% of cases noted to have metaplasia had diagnostic disagreement compared with 58% of cases not noted to have metaplasia, $p = 0.083$).

Discussion

With 2,601 specimens reviewed, to our knowledge, this study is the largest blinded review of the reproducibility of the 1994 WHO endometrial hyperplasia classification scheme to date. Our overall agreement ($\kappa = 0.71$), was similar to other studies which have κ values ranging from 0.2-0.7, Table 7.^{1, 10, 19, 24} Our agreement was lowest for the diagnosis of simple hyperplasia ($\kappa = 0.16$) and highest for “not hyperplastic” endometrium ($\kappa = 0.76$). Because of the high percentage of specimens with a final diagnosis of “no hyperplasia,” our overall agreement was inflated by our high agreement for non-hyperplastic specimens. In fact when “no hyperplasia” cases were excluded our overall agreement was significantly lower ($\kappa = 0.36$, specific agreement 80.6%). However, one might argue that having high numbers of non-hyperplastic specimens admixed with a variety of hyperplastic specimens is more analogous to actual clinical practice settings.

Although previous studies are not directly comparable due to differences in initial case selection and review panels, our agreement for specific diagnostic categories was similar both Zaino's 2006 prospective GOG study and Bergeron's 1999 European multi-institution study for the categories of non-atypical hyperplasia, atypical hyperplasia and adenocarcinoma. (Table 6) With κ values for agreement on atypical hyperplasia, a key diagnostic clinical decision point, as low 0.2-0.3 for “expert” gynecologic pathologists, there is clearly an issue with the reproducibility of the current WHO diagnostic scheme.

With specialists having so much trouble agreeing, it is not surprising that there is frequent disagreement between specialists and community pathologists. In Zaino's GOG study the majority review panel diagnosis supported the referring institution diagnosis in only 38% of cases submitted as atypical hyperplasia.²⁴ In our study, the final review panel diagnosis and the initial outside diagnostic categorization were not directly comparable, because the method of initial diagnosis categorization was intended to maximize the number of possible cases of complex and atypical hyperplasia selected for review (see Methods). Given our bias towards categorization of the initial diagnosis to a higher grade diagnostic category, it was not surprising that we had trends toward down-grading the initial diagnostic category by the review panel final diagnosis (data not shown). However, for the above reasons, agreement with the “original diagnosis” was not considered of value in this study.

What are the factors that cause diagnostic disagreement? Our study is the first of its kind to systematically investigate the contribution of sample adequacy, interpretation of key histologic features and the presence of complicating features (polyps and metaplasias) to diagnostic disagreement.

Clearly, problems with adequate sampling are an issue beyond effecting diagnostic agreement, with other studies showing the rates of finding “concurrent” carcinoma in hysterectomy specimens with a review panel diagnosis of normal or non-atypical hyperplasia as high as 19%.²⁴ But, with a given amount of diagnostic tissue, how do the amount of total tissue present for evaluation and the total amount of *hyperplastic* tissue present effect diagnostic agreement? In our study, specimens categorized by either pathologist as “scant” (< 0.5 cc) or “low volume” of hyperplastic tissue were significantly more likely to have disagreement about the diagnosis ($p = < 0.013$ and $p = < 0.0001$ respectively). This implies that specimens that have either a minimal amount of diagnostic tissue total (sample borders on inadequate), or samples that have only a very focal amount of hyperplastic tissue in otherwise normal endometrium, should be reviewed with caution. These samples may warrant a comment about the small amount of

diagnostic tissue present and the uncertainty in the diagnosis and request additional sampling. Zaino et al suggest similar findings related to sample adequacy, with greater diagnostic reproducibility for dilation and curettage specimens than office biopsy or curettage methods.²⁴ Interestingly, we did not find a statistically significant difference in diagnostic agreement between hysterectomy and non-hysterectomy specimens, possibly because even hysterectomies can have very low volumes of hyperplastic tissue, which can decrease diagnostic agreement.

While sampling is an issue that can be controlled by recommending additional tissue, the lack of objectivity in applying multiple diagnostic criteria to establish a diagnosis is more challenging. The histologic features referred to in the WHO as useful in establishing a diagnosis include “architectural changes,” “shift in the gland to stroma ratio,” and “cytologic atypia.”²¹ However, strict definitions of these features and the criteria used to establish a specific WHO diagnosis are not spelled out in great detail. Architectural changes said to be characteristic of complex hyperplasia include “irregular epithelial budding” and “increased gland complexity,” which is not further defined. A “shift in gland to stroma ratio in favor of the glands” is noted by the WHO to be a feature of complex hyperplasia as well but a strict threshold is not set. The endometrial intraepithelial neoplasia (EIN) scheme used by George Mutter and colleagues is more specific, using a volume percent stroma of less than 55% (area of glands > stroma) as one of the diagnostic criteria for a diagnosis of EIN.¹⁵ But it is the subjective interpretation of the presence of cytologic “atypia” in the WHO scheme that appears to be most problematic. In fact, the WHO specifically states that “definitions of cytologic atypia are difficult to apply in the endometrium because nuclear cytological changes occur frequently in hormonal imbalance, benign regeneration and metaplasia.”²¹ The WHO describes nuclear rounding, loss of polarity, prominent nucleoli, irregular nuclear membranes and cleared or dense chromatin as features of cytologic atypia but acknowledges that atypia may be best observed by comparison with the adjacent normal glands. The EIN scheme avoids using a descriptive definition of cytologic atypia and instead uses distinct cytology in the architecturally crowded focus that is different from background.¹⁵ Given the fairly loosely defined WHO diagnostic criteria, we were interested in determining if disagreement about the presence of key histologic features was a major factor in whether there was agreement about a specific WHO diagnosis.

Other studies have evaluated which histologic features could most aid recognition of cytologic atypia or architectural complexity. Kendall et al found “gland crowding” significantly associated with a diagnosis of complex hyperplasia while nucleoli was the only feature significantly associated with a pathologist calling a case atypical.¹⁰ Bergeron et al also found the presence of “gland crowding” most significantly associated with a diagnosis of hyperplasia while “nuclear pleomorphism” was most significantly associated with classification as atypical.¹ However, others have not investigated how concordance on the presence of certain histologic features specifically effect diagnostic agreement. Because our study did not include outcomes, we did not intend to define which features were more predictive of risk of carcinoma, but merely to investigate if we could agree on the presence of defined features and if disagreement of their presence effected agreement on final diagnosis.

In our study, cases with diagnostic disagreement were also more likely to disagree on specific key histologic features such as “architectural complexity,” “glandular crowding” and “cytologic atypia,” than cases with diagnostic agreement, indicating variable application of “defined” histologic features to formulate a diagnosis. Cytologic atypia was the feature most often disagreed on and had the largest difference between cases with disagreement versus agreement (47.3% of cases with diagnostic disagreement also disagreed on the presence of cytologic atypia, verses 16.1% of cases with diagnostic agreement). While our study only reflects the agreement between two pathologists, the poor reproducibility of atypical

hyperplasia in previous studies supports these findings.^{1, 24} Given that the presence of atypia is currently considered the best predictor of outcome in the WHO scheme, this finding calls into question the reliability of using “atypia” as it is currently defined (and variably interpreted) as a breakpoint for diagnostic categories.

The final factors we investigated as possible causes of diagnostic disagreement were “complicating” histologic features - the presence of features of a polyp or metaplasia. Because polyps tend to be less hormone responsive they can have more irregularly distributed glands with areas of crowding and have various metaplastic cytologic changes that can make differentiation of “normal” polyp from a polyp with areas of hyperplasia challenging. We did find a significant association with diagnostic disagreement in specimens where either pathologist had noted there were features of a polyp present (N = 230, $p < 0.0001$). Better criteria are needed to distinguish changes in polyps that should be considered higher risk, neoplastic lesions. In addition, when crowding or cytologic changes are limited to a polyp, a comment as to the unclear significance of the changes may be warranted.

The presence of metaplastic changes, or “epithelial cytoplasmic change”, in the endometrium varies from squamous, to tubal, to repair-associated eosinophilic syncytial change. The presence of extensive metaplasia can complicate the diagnosis of hyperplasia by making glands look more crowded (especially in extensive squamous metaplasia) or cytologically atypical. To complicate matters further, metaplastic changes are often associated with hyperplasias. We did note greater diagnostic disagreement when either pathologist noted the presence of metaplasia, however this did not reach statistical significance ($p=0.083$). This may have occurred because noting the presence of metaplasia was an optional part of the scoring form. In fact, of the 10 cases called “cannot rule out hyperplasia,” the most common reason noted was extensive metaplastic changes. Metaplastic changes are histologic features to be aware of as a possible pitfall in diagnosing endometrial hyperplasia but, while metaplasia was commonly associated with a diagnosis of “cannot rule out hyperplasia”, it was not a major cause of diagnostic disagreement in this study.

Additional studies to establish more reproducible criteria for endometrial hyperplasia that are also predictive of progression to carcinoma are needed. Various alternate diagnostic schemes have been proposed. Bergeron et al proposed combining simple and complex hyperplasia into a single “hyperplasia” group and combining atypical hyperplasia with a subset of well-differentiated carcinomas into an “endometrial neoplasia” group.¹ This scheme has the advantage of combining lesions that are treated in a similar way but its diagnostic utility has not been investigated. George Mutter and colleagues have more thoroughly investigated another scheme that was developed from molecular, histomorphometric and outcome data which separates pre-cancerous neoplastic “endometrial intra-epithelial neoplasia” from “benign endometrial hyperplasia” due to the presumed influence of unopposed estrogens.^{7, 13-17} However, while this system has the advantage of strong correlation of EIN with clonal populations, it is still unclear if this broad category can be further refined into high risk neoplasms that are likely to persist or progress to invasive carcinomas despite treatment with progestins versus lower risk neoplastic populations that may be spontaneously shed or regress with progestin therapy.

In conclusion, our study, the largest of its kind to date, confirms previous findings related to the poor reproducibility of the current WHO endometrial hyperplasia classification system. In addition, our findings suggest that diagnostic disagreement is due both to inability to agree on the presence of various key histologic features and the amount of diagnostic tissue present. We suggest that in the clinical setting specimens with limited amounts of diagnostic tissue (either low volumes of diagnostic hyperplastic tissue or overall scant specimens) should be interpreted with caution and that recommending additional tissue should be considered. Setting a threshold

for amount of diagnostic tissue present necessary for a definitive diagnosis could improve diagnostic agreement and decrease the rates of immediate hysterectomies for this usually low-grade neoplastic process. Given the poor reproducibility of the diagnosis of endometrial hyperplasia, studies examining outcomes in this field may also want to consider limiting their cases to those that have diagnostic agreement among reviewing pathologists or at least have a minimum threshold of diagnostic tissue present. Using stricter criteria for outcome studies will help give a clearer picture of the natural history of endometrial hyperplasia and perhaps shed more light on which lesions are truly higher risk.

Appendix

Appendix 1

Scoring of hyperplasia cases

Adequacy of Sample:

- Scant endometrial tissue (<0.5 cc)
- Moderate endometrial tissue (> 0.5 cc, one block)
- Abundant endometrial tissue (> one block)

Volume of Hyperplasia:

- Low (<0.5 cc)
- High (>0.5 cc)

Degree of Glandular Crowding:

- Low (gland:stroma ratio > 50:50 but < 80:20)
- High (gland:stroma ratio > 80:20)

Architectural Complexity:

- Low (Focal architectural complexity without marked branching, cribriforming, or back to back glands)
- High (Severe architectural complexity - marked branching, cribriforming or back to back glands)

Cytologic Atypia:

- Absent (No significant cytologic atypia; similar to "normal" endometrium in same specimen)
- Present (Significant cytologic atypia — rounding of nuclei, loss of polarity, prominent nucleoli, nuclear pleomorphism)

Metaplasia:

☐ Present

Features of Polyp:

☐ Present

References

1. 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. [PubMed: 10478671]
2. Beutler HK, Dockerty MB, Randall LM. Precancerous lesions of the endometrium. *Am J Obstet Gynecol* 1963;86:433–43. [PubMed: 13967897]
3. Campbell PE, Barter RA. The significance of a typical endometrial hyperplasia. *J Obstet Gynaecol Br Commonw* 1961;68:668–72. [PubMed: 13690268]
4. Fleiss, J. Statistical methods for rates and proportions. Vol. second edition. John Wiley & Sons; New York: 1981. p. 212-236.

5. Gore H, Hertig AT. Carcinoma in situ of the endometrium. *Am J Obstet Gynecol* 1966;94:134–55. [PubMed: 5321298]
6. Gusberg SB, Kaplan AL. Precursors of Corpus Cancer. Iv. Adenomatous Hyperplasia as Stage O Carcinoma of the Endometrium. *Am J Obstet Gynecol* 1963;87:662–78. [PubMed: 14074031]
7. Hecht JL, Ince TA, Baak JP, Baker HE, Ogden MW, Mutter GL. Prediction of endometrial carcinoma by subjective endometrial intraepithelial neoplasia diagnosis. *Mod Pathol* 2005;18:324–30. [PubMed: 15529181]
8. Hendrickson, MKR. *Surgical Pathology of the Uterine Corpus*. WB Saunders Co.; Philadelphia: 1980.
9. Horn LCSU, Bilek K, Hentschel B, Eienkel J. Risk of progression in complex and atypical endometrial hyperplasia: clinicopathologic analysis in cases with and without progestin treatment. *International Journal of Gynecological Cancer* 2004;14:348–353. [PubMed: 15086736]
10. 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;22:1012–9. [PubMed: 9706982]
11. 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. [PubMed: 4005805]
12. Landis JRKG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74. [PubMed: 843571]
13. Mutter GL. Endometrial intraepithelial neoplasia (EIN): will it bring order to chaos? The Endometrial Collaborative Group. *Gynecol Oncol* 2000;76:287–90. [PubMed: 10684697]
14. Mutter GL. Histopathology of genetically defined endometrial precancers. *Int J Gynecol Pathol* 2000;19:301–9. [PubMed: 11109157]
15. 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;190:462–9. [PubMed: 10699996]
16. Mutter GL, Ince TA, Baak JP, Kust GA, Zhou XP, Eng C. Molecular identification of latent precancers in histologically normal endometrium. *Cancer Res* 2001;61:4311–4. [PubMed: 11389050]
17. Mutter GL, Zaino RJ, Baak JP, Bentley RC, Robboy SJ. Benign endometrial hyperplasia sequence and endometrial intraepithelial neoplasia. *Int J Gynecol Pathol* 2007;26:103–14. [PubMed: 17413975]
18. Pecorelli S, Benedet JL, Creasman WT, Shepherd JH. FIGO staging of gynecologic cancer. 1994-1997 FIGO Committee on Gynecologic Oncology. International Federation of Gynecology and Obstetrics. *Int J Gynaecol Obstet* 1999;65:243–9. [PubMed: 10428343]
19. 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;16:33–7. [PubMed: 8986530]
20. Tavassoli F, Kraus FT. Endometrial lesions in uteri resected for atypical endometrial hyperplasia. *Am J Clin Pathol* 1978;70:770–9. [PubMed: 717282]
21. Tavassoli, FADPE., editor. *World Health Organization Classification of Tumors: Pathology and Genetics of Tumours of the Breast and Female Genital Organs*. IARC Press; Lyon: 2003.
22. Trimble CL, Kauderer J, Zaino R, et al. Concurrent endometrial carcinoma in women with a biopsy diagnosis of atypical endometrial hyperplasia: a Gynecologic Oncology Group study. *Cancer* 2006;106:812–9. [PubMed: 16400639]
23. Vellios F. Endometrial hyperplasias, precursors of endometrial carcinoma. *Pathol Annu* 1972;7:201–29. [PubMed: 4557932]
24. Zaino RJ, Kauderer J, Trimble CL, et al. Reproducibility of the diagnosis of atypical endometrial hyperplasia: a Gynecologic Oncology Group study. *Cancer* 2006;106:804–11. [PubMed: 16400640]

Figure 1A

Figure 1B

Figure 1A: Diagnostic trends of the two primary panel pathologists for all specimens. Pathologist A was more likely to diagnose atypical hyperplasia or carcinoma while pathologist B was more likely to diagnose complex, simple or no hyperplasia. However, overall diagnostic trends were similar.

Figure 1B: Diagnostic trends for cases sent to third reviewer. The third reviewer, C, agreed with 43.9% of pathologist B's and 27.7% of pathologist A's diagnoses. The third reviewer diagnosed atypical hyperplasia more often than reviewer B (but less than A), "no hyperplasia" or simple hyperplasia more frequently and complex hyperplasia less frequently than both reviewers A and B. Reviewer C also had a similar frequency of diagnosing carcinoma to reviewer A.

Figure 2.

Examples of the histology of cases with agreement and disagreement on the presence of cytologic atypia. **A)** Case with agreement on both the presence of cytologic atypia and the diagnosis of atypical hyperplasia. **B)** Case that was quantified as “scant” and “low volume neoplasia” with disagreement about the presence of cytologic atypia. The case had a final classification as atypical hyperplasia. **C)** Case with disagreement on the presence of cytologic atypia, which had a final classification as complex hyperplasia without atypia. The most atypical appearing focus is pictured, which was very focal. Original magnification $\times 200$ (A-D).

Table 1

Clinical characteristics of cohort

	N	%
<i>Age (years)</i>		
<i><45</i>	288	16.8
<i>45-54</i>	752	43.9
<i>55-64</i>	379	22.1
<i>≥65</i>	293	17.1
<i>Race</i>		
<i>White</i>	1455	85.0
<i>Black</i>	44	2.6
<i>Asian</i>	75	4.4
<i>Other</i>	29	1.7
<i>Unknown</i>	109	6.4
<i>Body mass index (kg/m²) at reference date</i>		
<i>Not overweight (<24.9)</i>	531	32.9
<i>Overweight (25-29.9)</i>	406	25.2
<i>Obese (30 -39.9)</i>	426	26.4
<i>Morbidly obese (≥40)</i>	251	15.6
<i>Missing</i>	98	

Table 2

Frequencies of WHO final panel diagnoses

Final Diagnosis	Frequency	Percent
<i>No hyperplasia or simple hyperplasia</i>	1829	70.3%
<i>Complex hyperplasia</i>	396	15.2%
<i>Atypical hyperplasia</i>	288	11.1%
<i>Adenocarcinoma</i>	54	2.08%
<i>Other</i> [*]	34	1.31%
<i>Total</i>	2601	

^{*} Inadequate specimen, non-diagnostic or cannot rule out hyperplasia

Table 3

Diagnostic agreement by diagnosis

	Percent specific agreement	Kappa
<i>No hyperplasia</i> <i>N agreed = 1370</i>	90.3%	0.76
<i>Simple hyperplasia</i> <i>N agreed = 67</i>	31.1%	0.16
<i>Complex hyperplasia</i> <i>N agreed = 226</i>	51.1%	0.21
<i>Atypical hyperplasia</i> <i>N agreed = 139</i>	49.8%	0.35
<i>Adenocarcinoma</i> <i>N agreed = 44</i>	57.5%	0.55
<i>Overall except "no hyperplasia"</i> <i>N total = 1,161</i>	80.6% [*]	0.36 [*]
<i>Overall</i> <i>N total = 2,531</i>	91.1% [*]	0.71 [*]

^{*} Weighted percent agreement or weighted kappa

Table 4

Correlation of quantity of diagnostic material with diagnostic disagreement

	Cases with diagnostic disagreement	
	N	%
<i>Specimen Adequacy</i>		
<i>"Scant" (< 0.5cc)</i> <i>N = 353</i>	228	64.6
<i>Not scant N= 791</i>	447	56.5
P value of difference		0.013
<i>Volume of Hyperplasia</i>		
<i>Low volume (<0.5 cc)</i> <i>N= 761</i>	471	61.9
<i>High volume (>0.5cc)</i> <i>N = 227</i>	88	38.8
P value of difference		<.0001
<i>Type of Specimen</i>		
<i>Biopsy or curettage</i> <i>N = 1060</i>	629	59.3
<i>Hysterectomy</i> <i>N = 101</i>	46	55.5
P value of difference		0.51

Table 5
Histologic features involved in diagnostic disagreement

	Disagreed on cytologic atypia		Disagreed on degree of glandular crowding		Disagreed on architectural complexity	
	N	%	N	%	N	%
<i>Cases with diagnostic agreement N= 417</i>	67	16.1%	88	21.1%	92	22.1%
<i>Cases with diagnostic disagreement N= 364</i>	172	47.3%	123	33.5%	115	31.3%
P value	<.0001		<.0001		.005	

Table 6
Comparison of studies on 1994 WHO endometrial hyperplasia classification scheme reproducibility

Study	N total	Case selection	Reviewers	Overall kappa	NH	SH	CH	AH	CA
Shov 1997	128	Review of cases submitted as hyperplasia	Two institutions in Denmark, 6 pathologists	0.20-0.25	NA	0.21-0.25	0.07-0.15	0.5-0.65	NA
Kendall 1998	100	Review of a selected variety of diagnoses	Single US institution, 5 pathologists	0.7	0.9	0.6		0.5	0.8
Bergeron 1999	56	Review of a selected variety of diagnoses	Multiple European institutions, 5 pathologists	0.47	71%	44%	18%	25%	49%
Zaino 2006	302	Review of cases submitted as atypical	Multiple US hospitals, 3 pathologists	0.4	0.48	0.38		0.28	0.51
Allison 2007	2,601 (1,161 "not normal")	Women with a possible diagnosis of hyperplasia	Single institution, 3 pathologists	0.71 91% 0.36 (for hyperplasia only)	0.76 90%	0.16 31%	0.21 51%	0.35 50%	0.55 58%

NH = no hyperplasia, SH = simple hyperplasia, CH = complex hyperplasia, AH = atypical hyperplasia, CA = carcinoma
% = % agreement

Kappa: 0.0-0.2=slight agreement, 0.21-0.40=fair, 0.41-0.60=moderate, 0.61-0.80=substantial, 0.8-1.00=almost perfect agreement 16