



Published in final edited form as:

Histopathology. 2016 December ; 69(6): 1028–1046. doi:10.1111/his.13035.

Histologic Features associated with Diagnostic Agreement in Atypical Ductal Hyperplasia of the Breast: Illustrative Cases from the B-Path Study

Kimberly H. Allison¹, Mara H. Rendi², Sue Peacock³, Tom Morgan³, Joann G. Elmore³, and Donald L. Weaver⁴

¹Department of Pathology, Stanford University School of Medicine, Stanford, CA

²Department of Pathology, University of Washington School of Medicine, Seattle, WA

³Department of Medicine, University of Washington School of Medicine, Seattle, WA

⁴Department of Pathology and University of Vermont Cancer Center, University of Vermont, Burlington, VT

Abstract

Background—Case specific characteristics associated with interobserver diagnostic agreement in atypical ductal hyperplasia (ADH) of the breast are poorly understood.

Methods—Seventy-two test set cases with a consensus diagnosis of ADH from the B-Path study were evaluated. Cases were scored for 17 histologic features which were then correlated with the participant agreement with the consensus ADH diagnosis.

Participating pathologists' perceptions of case difficulty, borderline features, or if they would obtain a second opinion were also examined for associations with agreement.

Results—Of the 2,070 participant interpretations on the 72 consensus ADH cases, 48% were scored by participants as difficult and 45% as borderline between two diagnoses; the presence of both of these features was significantly associated with increased agreement ($p < 0.001$). A second opinion would have been obtained in 80% of interpretations, and this was associated with increased agreement ($p < 0.001$). Diagnostic agreement ranged from 10–89% on a case-by-case basis. Cases with papillary lesions, cribriform architecture and obvious cytologic monotony were associated with higher agreement. Lower agreement rates were associated with solid or micro-papillary architecture, borderline cytologic monotony or cases without a diagnostic area that was obvious on low power.

Conclusions—The results of this study suggest that pathologists frequently recognize the challenge of ADH cases with some cases more prone to diagnostic variability. In addition, there are specific histologic features associated with diagnostic agreement on ADH cases. Multiple example images from cases in this test set are provided to serve as educational illustrations of these challenges.

Keywords

Breast Pathology Study (B-Path); diagnostic agreement; atypical ductal hyperplasia; breast pathology; interobserver agreement

Introduction

In most areas of pathology, diagnostic agreement is high. However, significant interobserver diagnostic disagreement exists in areas that are considered subtle biologic “grey zones.”^{1–15} In breast pathology, a particularly problematic area is the distinction between atypical ductal hyperplasia (ADH) and limited extent low-grade ductal carcinoma in situ (DCIS); lesions that exist in a biologic continuum but that have been separated artefactual on the basis of often subtle histologic differences and extent.^{16–28} Because major clinical treatment thresholds exist between a diagnosis of ADH and DCIS, the high diagnostic variability for these two entities has been a focus of negative media attention suggesting that pathologists are “prone to error,” and that pathologists “frequently misdiagnose breast tissue.”^{18,21,25,26,29–32} However, the reasons underlying diagnostic variability in this area are complex and can vary from subtle differences of professional opinion on borderline cases, to missing focal findings that may have marginal clinical significance.³³ An improved understanding, by pathologists and the community at large, of the reasons underlying diagnostic variability may help shed light on ways to improve concordance and improve communication regarding the borderline character of certain cases.

The B-Path study examined diagnostic variability for breast tissue test set samples evaluated by a panel of specialized breast pathologists and 115 participating pathologists.³⁴ This study reported very high (96%) agreement with the expert consensus diagnoses for invasive carcinoma, and somewhat higher diagnostic variability for DCIS (84%) and benign cases without atypia (87%). However, agreement was 48% for the ADH category, which included both ADH and intraductal papilloma with ADH (IPA). There were statistically significant associations between lower concordance with the expert consensus diagnosis and pathologists who reported lower weekly breast case volumes or who practice in nonacademic settings. These associations suggest that educational interventions or training sets may improve diagnostic agreement in this challenging area.

The 240 test set cases from the B-Path study are a uniquely well-characterized set of slides that can be used to examine features associated with diagnostic agreement and disagreement in breast pathology. By focusing on the most problematic cases, the 72 categorized as ADH by the expert consensus panel, a detailed evaluation of the specific histologic features present may highlight the most diagnostic and problematic features. In this study, we reviewed the 72 consensus ADH slides, scored a set of features ranging from low magnification to high magnification findings and evaluated associations with low versus high diagnostic agreement on a case by case basis. We also present a series of images from these B-Path test set cases to help illustrate the histologic features associated with diagnostic agreement on ADH.

Methods

Case Selection and Participant Analysis of Cases

The B-Path study case selection methodology and establishment of consensus diagnoses is detailed elsewhere.^{35,34} The 72 cases from the B-Path study with an expert consensus diagnosis of ADH or ADH involving a papilloma were selected for review and additional analysis for this study. Participants reviewed a single glass slide per case from one of four test sets each comprised of 60 cases; digital whole slide image analysis results were not included in this study. Each test set contained a range of cases from benign without atypia to invasive carcinoma and were matched for difficulty and categorical diagnosis. There were 2,070 independent participant diagnoses for the 72 ADH cases.^{34,36} For each case, participants also scored the level of diagnostic difficulty (on a Likert scale from 1–6 with 1 being very easy and 6 as very challenging), how confident they were in their diagnosis (on a Likert scale from 1–6 with 1 being very confident and 6 not at all confident), if they would desire a second opinion on the case (either because it was their practice’s policy to obtain or because they desired one or both) and if it was borderline between two diagnoses. For the cases considered borderline, participants indicated the two diagnoses they considered in addition to their final selection.

Analysis of Histologic Features on Atypia Cases

All 72 consensus atypia cases used in the B-Path study were reviewed by three breast pathologists (KHA, MHR, DLW) in the archived digital whole slide image (WSI) format to further analyze the characteristics present. After discussion of potentially relevant features, a set of 17 histologic features was included to be scored. These features are shown in the Figure 1 example of the histologic features score sheet).

Scored features typically associated with initial screening of a case included the following: 1) Number of regions of potential interest to screen (When the slide is viewed on low power, how many areas are you interested in seeing on higher power?) 2) Was the diagnostic area obvious on low power? (Was the area diagnostic of ADH considered easily recognizable on low power?) 3) Were distracting diagnoses present in other areas? (Were other diagnoses present that could have potentially distracted a reviewing pathologist from the diagnosis of ADH and if so what were they?) 4) Were other diagnoses in the differential or present within the same lesion that contained ADH, and if so what were they?

Scored features that are more apparent on low power or related to extent were as follows: 1) Is the lesion of interest (ADH-containing lesion) a papillary lesion? 2) What are the number of separate areas with atypia/lesion containing atypia? (“Area” was defined as a contiguously involved region of the slide by the intraductal lesion containing ADH—but the area could include other associated diagnoses, such as FEA, as well.) 3) Largest single area with atypia/lesion containing atypia in centimeters. For example, a slide with a 3 mm area of predominantly FEA with 2–3 scattered membrane-bound spaces containing ADH, would be sized based on the size of the entire lesion as 3 mm. 4) Number of foci (defined as membrane-bound spaces/glands) involved by ADH. 5) Largest single discrete focus (membrane bound space) involved by ADH (2 mm or > 2 mm).

Scored features that typically require more detailed higher, power examination included analysis of the cytology and the architecture of the lesions containing ADH. The following cytologic features were scored: 1) Cytologic monotony of lesion classified as ADH (very monotonous with uniformly rounded nuclei vs questionably monotonous) 2) Apocrine cytology, 3) Nuclear hyperchromasia, 4) Prominence of nucleoli and 5) Presence of calcifications. The following architectural features were scored: 1) Predominant type of architecture present in the lesion containing ADH (Cribriform = Growth into structures that form spaces that are often polarized around the lumens. Micropapillary = Club or finger-shaped projections of epithelium. Solid = Filling a duct or acinus in a uniform architectural pattern without formation of obvious cribriform lumens/spaces. Papillary = Epithelial growth around intra-ductal fibrovascular cores. Flat = Predominantly flat, dilated duct spaces with only subtle early arches and bridges.) The development of the architecture was scored as one of the following categories: 1) Well-developed atypical architecture (polarized cribriform spaces/microacini, club shaped structures, etc typical of low grade DCIS) 2) Partially developed architecture (incomplete cribriforming, arches, bridges, limited micropapillae), 3) Solid architecture, and 4) No or subtle architectural atypia (limited architecture or architecture streaming/more typical of usual hyperplasia). Lastly, the uniformity of the combined cytologic and architectural features diagnostic of ADH within the lesion of interest was scored as either uniformly involved or mixed.

The algorithm to determine the final score for each feature for each case used a conditional method. First, a score was considered final if two of the experts (KA and MR) agreed on the histologic feature for a particular case. Overall, this occurred in 97% of all histologic features reviewed across all cases. If there was disagreement, then the score of a third pathologist (DW) was used and the final score selected was the score where two of the three pathologists agreed. For a small number of features in rare cases (8/1224 feature assessments, <1%), there was no agreement among the three reference pathologists and the case was not included in the overall scoring for that particular histologic feature.

Data Analysis

For comparisons between the participant diagnosis and the consensus reference diagnosis on each case, the highest order diagnosis was used to assign a final case diagnosis within the following categories: 1) Benign (including proliferative lesions like columnar cell hyperplasia and usual ductal hyperplasia as well as non-proliferative breast tissue), 2) flat epithelial atypia (FEA) and lobular neoplasia (LN; atypical lobular hyperplasia and lobular carcinoma in situ), 2) ADH (including both ADH and intraductal papilloma with ADH), 3) DCIS, and 4) Invasive carcinoma.

The participant diagnoses on all 72 cases were dichotomized based on agreement with the consensus reference diagnosis of ADH. Associations were assessed between participant agreement with the consensus ADH diagnosis and participant assessments of the case's challenges (level of difficulty, case borderline between two diagnoses, would request a second opinion, and confidence in their diagnosis). Summary statistics including frequencies and percents were produced. P-values were computed using generalized estimating equation (GEE) models accounting for repeated measures.

Comparisons were made to determine whether the 17 histologic features varied across agreement using generalized estimating equation (GEE) models accounting for repeated measures with a binomial distribution for the dependent variable. The histologic features were also examined for associations by the specific participant diagnosis. Features found to be independently significant at $P < 0.05$ with agreement were further analyzed by multivariate regression analysis again using GEE accounting for repeated measures with a binomial distribution. All analyses were performed using SAS software, version 9.4.

Results

Participant ratings of case challenges in ADH test set cases and associations with agreement

Table 1 summarizes data on participants' ratings of the case difficulty level, if they considered the case borderline between two diagnoses, if they would have requested a second opinion and their confidence level in their interpretation. The 72 consensus ADH cases had a high frequency of interpretations of the cases as difficult (48% of interpretations) or borderline (45% of interpretations) by participants. For 80% of case interpretations, participants indicated they would have requested a second opinion. Interestingly, case interpretations considered difficult, borderline or needing a second opinion by participants had a significantly higher percent agreement with the consensus ADH diagnosis than cases not considered to have these challenges ($p < 0.001$). Case interpretations not scored as difficult or borderline and those where a second opinion would not have been obtained were significantly more likely to have been called Benign/FEA/LN by participants ($p < 0.001$). Most participants recorded high confidence levels in their interpretations (72%), despite the high frequency of interpretations considered difficult, borderline or needing a second opinion.

Range of diagnostic agreement and alternative diagnoses on consensus ADH cases

The diagnostic agreement on the 72 cases with a consensus diagnosis of atypia ranged by case from 10–89% (Figure 2). Agreement with the consensus diagnosis of ADH occurred for 48% of case interpretations with the remainder of interpretations quantified as follows: 25% benign, 10% FEA/LN, 17% DCIS and $< 0.10\%$ invasive carcinoma.

On a case-by-case basis for these consensus ADH cases, the frequency of a diagnosis other than ADH ranged as follows: DCIS diagnosis ranged from 0–57% 0–70% for benign diagnoses, and 0–62% for FEA/LN diagnoses. Of the non-ADH FEA/LN diagnoses, 43% (90/207 interpretations) had a primary (and highest order) diagnosis of FEA and 57% LN (117/207 interpretations).

Association of histologic features scored with agreement on ADH and alternative diagnoses

Of the 17 histologic features evaluated, only 6 features were statistically significantly associated with agreement with the consensus ADH diagnosis. These features, and the others evaluated, are listed in Table 2 in order of most significant associations with diagnostic agreement that the case was ADH in univariate analysis. Multivariate regression

analysis was performed including all features found significantly associated with agreement with the expert consensus individually. Architectural pattern present in the lesion was the most significant feature in predicting agreement with the consensus ADH diagnosis ($P < 0.001$), followed by cytologic monotony presence/absence ($p = 0.032$). Table 3 shows the distribution of the participant diagnoses for each histologic feature.

Example images from the two cases with the highest agreement (83% and 89% of participants agreed it was ADH) are shown in Figure 3. These two cases both had a single area of involvement that was < 2 mm, the area of interest was obvious on low power, had obvious cytologic monotony and cribriform architecture.

Papillary lesions ($N = 14$) had the highest agreement rates (56%) vs non-papillary lesions (46%, $N = 58$) ($p < 0.001$). However, in multivariate regression analysis it no longer maintained a significant association with agreement ($p = 0.11$). Papillary lesions were less frequently called Benign (23%) than non-papillary lesions (25%) and there was less confusion with FEA/LN (only 1% of papillary lesions were called FEA/LN versus 12% of non-papillary lesions). The most common benign diagnosis on these cases was intraductal papilloma without atypia. Figure 4 shows images from two of the papillary cases with higher agreement on the diagnosis of ADH/IPA.

The vast majority of cases had cribriform architecture and this pattern was also associated with higher agreement (51%). Figure 5 shows images from a case with cribriform architecture that the majority of participants called ADH.

Cases with solid or micropapillary architecture had very poor agreement with the consensus diagnosis of ADH (27% and 40% respectively, $p < 0.001$).

Cases with solid architecture were frequently (37%) diagnosed as FEA/LN or DCIS (26%). Figure 6 shows images from a case with solid architecture that only 17% called ADH, 28% called DCIS and the majority (55%) called LN (either ALH or LCIS). Subtle microacinar structures were present in this case supporting a ductal phenotype. This differential is often resolved with E-cadherin immunohistochemistry which was not available to participants.

Cases with micropapillary pattern architecture were also frequently diagnosed as Benign (21%), FEA/LN (13%) or DCIS (26%). Cases frequently diagnosed as FEA or columnar cell changes tended to have very focal micropapillary findings. Figure 7 shows two examples of cases with micropapillary architecture that were most frequently called Benign or FEA/LN by participants. The most frequent specific diagnosis in this category was FEA or columnar cell hyperplasia (100%) for the case in Panel B. The case in Panel A had a large adjacent area of LCIS so the majority of participants whose diagnoses were in the FEA/LN category had called the case LN, perhaps not noting the area pictured. 26% of micropapillary pattern cases were diagnosed as DCIS. These cases tended to have more areas/foci with the lesion of interest. Only three cases were classified as “flat pattern” ADH, with predominantly flat architecture and subtle arches, bridges or early micropapillae. FEA was often in the differential diagnosis, with ADH architecture present only focally. Examples of this pattern are shown in Figure 10.

Of the cytologic features scored, cytologic monotony was significantly associated with agreement that a case was ADH. Cases that were scored as “very monotonous” had an overall agreement of 50%, versus 43% for cases scored as “not monotonous or borderline monotonous” ($p = 0.013$). Examples of cases with borderline monotony are shown in Figures 6, 8 and 9. Examples of cases scored as very monotonous are shown in Figures 3 and 5.

The specimen type was also significantly associated with agreement on ADH. Slides with a core biopsy sample had higher agreement than slides of an excision (50% for core biopsies, 44% for excisional biopsies; $p = 0.009$). Whether this was due to the larger volume of tissue to examine in an excision or due to differences in the clinical impact of an ADH diagnosis on a core vs excisional sample was not evaluated. However, this feature was less significant on multivariate regression analysis ($p = 0.18$). Agreement was also significantly higher for cases where the diagnostic area was considered obvious on low power (50% when obvious, 45% when not obvious; $p = 0.020$; $p = 0.31$ on multivariate regression analysis). The two cases with the highest agreement on an ADH diagnosis both had the area of interest obvious on low power (Figure 3).

The number of foci (defined as number of membrane bound spaces) involved by ADH (frequently this was partial/incomplete involvement of a membrane bound space by the proliferation) was significantly associated with agreement as well ($p = 0.041$), but not on multivariate regression analysis ($p = 0.21$). Increasing numbers of foci was most prominently associated with frequency of making a DCIS diagnosis on these cases with 7% of cases with 1–2 foci called DCIS compared with 28% of cases with 6+ foci called DCIS.

Discussion

This study provides an in depth look at the diagnostic challenges of ADH in breast pathology. Using a well-characterized test set of 72 cases that were classified as ADH by an expert panel consensus diagnosis and the detailed diagnostic scoring data from 115 participating pathologists evaluating these cases, the specific case-based factors associated with agreement that a case was ADH were evaluated. The results of this study suggest that; 1) pathologists frequently recognize the inherent challenge of ADH cases, 2) the diagnostic agreement varies dramatically by specific case, and 3) there are specific histologic features that are associated with diagnostic agreement on ADH cases. Using this information and illustrative case examples, the complexity of these cases can be better understood and potentially serve as consensus-building tools to improve diagnosis and management of these challenging lesions.

Participants recognized that the majority of cases were difficult or borderline with other diagnoses, and very frequently indicated that they would want a second opinion on a case (80% of interpretations). Interestingly, when these features were indicated, there was higher agreement with the expert consensus diagnosis of ADH. When they were not indicated, these cases were more frequently diagnosed as Benign. These data suggest that pathologists do recognize the diagnostic challenges and diagnostic variability of intra-ductal proliferative lesions like ADH, and would frequently utilize colleagues to help build diagnostic consensus

on these lesions. However, some of the more subtle features considered diagnostic for ADH by the consensus panel were less frequently recognized or considered ADH by participants. In addition, because the cases were screened independently by the three consensus panelists and then reviewed in consensus, there may have been a detection bias for rare or inconspicuous findings.

Other qualitative analysis from the B-Path expert consensus review meetings suggests that specialist breast pathologist's diagnostic variability is often due to subtle differences in professional opinion, most frequently in cases where the diagnosis of ADH is being considered.³³ The expert panel had an initial independent agreement of 80% on cases eventually classified as ADH following meetings to establish consensus diagnosis of the ADH cases.³⁴ Required second opinions on ADH cases are currently not common as pathology practice policy (only 36% of pathologists surveyed in the B-Path study indicated required review of ADH was policy at their institution), but in practice, second opinions may occur much more frequently (83.9% of those surveyed indicated they obtain second opinions in at least some of their ADH cases and 28.0% obtain one in all ADH cases in the absence of a policy requiring it).³⁷

Diagnostic variability for the ADH test set cases appeared to be case-based, with concordance ranging from as low as 10% to as high as 89% (overall agreement of 48%). As illustrated by Figure 2, some cases had high frequencies of Benign interpretations vs ADH, while others had high frequencies of other atypias (FEA/FEA) or DCIS interpretations vs ADH. However, in our study there were also cases that had high proportions of interpretations in all 4 of the most common categories (Benign, FEA/LN, ADH and DCIS), as the case featured in Figure 9.

Because these cases were not pre-selected for their challenging or illustrative features, they likely serve as a reasonable representation of the spectrum of cases seen in clinical practice in the ADH category. Given the wide spectrum of case-by-case variability in diagnostic agreement, evaluating the features of the cases with the lowest or highest agreement may be more illuminating than evaluating overall agreement rates for all cases. Some cases may have features that are not easily classified into a single category, but are considered borderline with another diagnosis. In practice the differential of these lesions often falls into one of these three categories: 1) ADH vs benign proliferations like UDH, 2) ADH from other atypias like FEA or LN, and 3) ADH vs low grade, limited extent DCIS.

By examining the histologic features present on a particular case in this test set and the frequency of agreement with the consensus ADH diagnosis of the cases, the specific challenges of this diagnosis can be analyzed in more depth. Higher agreement with an ADH diagnosis was observed for "classic" examples that were cribriform pattern with obvious cytologic monotony, but were limited in extent and focality. Papillary lesions with ADH were more also frequently identified as ADH by participants than non-papillary lesions. This may be due to several factors, such as papillary lesions generally being larger, and more obvious on low power. However, this study did not examine the frequency of non-atypical papillomas being called atypical, which would be necessary to draw significant conclusions about papillary lesions and diagnostic agreement. Papillary lesions have been noted to make

up a significant proportion of cases sent by pathologists for consultation by specialized breast pathologists.^{38,39}

The biggest challenge in the ADH spectrum was the low agreement for cases with borderline cytologic monotony, solid or micropapillary growth patterns. For cases with borderline cytologic monotony, the differential diagnosis can be either with a polyclonal process such as usual ductal hyperplasia (UDH), ADH or, on the other end of the spectrum, an intermediate grade DCIS. This differential diagnosis can often be addressed by additional immunohistochemical studies, which were not available to participants in this study. A CK5/6 antibody stain should stain a polyclonal in-situ process such as UDH in a mixed or “mosaic” pattern, while a neoplastic low/intermediate grade process should be uniformly CK5/6 negative (with the exception of the surrounding myoepithelial cells).^{40,41} Estrogen receptor (ER) should also stain a polyclonal, hyperplastic process like UDH in a mixed pattern, while a low/intermediate grade neoplasia should be uniformly and strongly ER positive (although lesions such as columnar cell lesions will also have this “neoplastic” pattern).⁴² However, because most columnar cell lesions are also CK5/6 negative and ER uniformly positive, these stains are not useful to distinguish between columnar cell lesions from ADH or DCIS. In addition, since both ADH and DCIS have identical staining patterns, IHC is also not useful in distinguishing between ADH and DCIS. However, some studies have demonstrated increases in diagnostic agreement when cocktail stains such as ADH-5 are used on intraductal proliferative lesions (ADH-5 stains includes CK5, 14, 7 18 and p63), with more cases clearly classified as UDH.⁴³ Because of the ability to resolve some of these cases with borderline cytologic monotony with IHC stains, agreement may be higher in practice for these lesions. However, pathologists should be aware of the differential and tools available to resolve this particular diagnostic dilemma.

For cases with a solid intra-ductal proliferative pattern, the alternative diagnosis to ADH was frequently a lobular in situ lesion (ALH or LCIS). This differential can also often be resolved with IHC stains. A lobular process will typically be E-cadherin negative while a ductal process will have membranous E-cadherin expression.⁴⁴ Although E-cadherin stains were not available to participants in this study, subtle features were considered diagnostic of ADH on H&E by the expert pathologist panel on these cases, such as the micro-acini or small polarized lumens formed by cells in cases such as those pictured in Figure 6. However, it is notable that currently there is little difference in the clinical management of ADH and ALH/LCIS lesions on excision biopsy (both of which are typically managed as risk lesions).

In this study, cases with a micropapillary growth pattern ranged from cases with predominantly FEA or columnar cell hyperplasia and focal formation of micropapilla to cases with multiple foci of a micropapillary process that bordered on micropapillary DCIS. For micropapillary processes, particularly on core biopsy, multiple levels or mounting ribbons can help clarify the extent of the process. Similar to non-papillary cases, the extent of the process can be critical in determining the diagnosis with certainty. If the changes are considered borderline between two of these diagnoses, or the process is not uniform or is close to a 2mm threshold, a conservative approach with the lesser diagnosis in the differential is recommended until more complete evaluation of the lesion can be performed on a surgical excision specimen. Sometimes CK5/6 staining can also help distinguish

micropapillary UDH from micropapillary ADH or DCIS, although staining patterns are often not as robust as in florid examples of UDH with a non-micropapillary pattern.

When examining data on diagnostic agreement in a test-set setting, it is important to emphasize the differences from actual clinical practice. In contrast to clinical practice, where many slides, levels or extra stains may be examined, participants in this study only reviewed a single H&E stained slide. This was considered the most practically feasible way to conduct this study which involved sending glass test set slides to be circulated between multiple participants. However, in an effort to mimic clinical practice and in contrast to several studies on diagnostic agreement of intraductal proliferative lesions, the areas of interest were *not* pre-selected/indicated on the slides.^{25,26,29} Participants were also not able to request second opinions, which, as discussed above, are an important component of real-world clinical practice (and were something participants could note they would have requested on each case). Lastly, in our study clinical information was not available to participants other than the patient's age and specimen type (core vs excisional biopsy). In clinical practice, additional information about the imaging findings, risk factors, prior relevant pathology, etc is frequently available either from medical records or discussions with clinicians and can help inform pathologist decision-making.

Perfect agreement for the spectrum of ADH lesions is unlikely to be achievable, even with well-established criteria. Many factors are influential when making a diagnosis in this biologic "grey zone." This is inherently challenging because of the lack of clear biologic distinction between lesions that exist in a continuum. The clinical context is likely taken into consideration when making these diagnoses, a factor that was not possible beyond the type of specimen being analyzed in this test-set based study.³³ The fact that diagnostic agreement was higher in core biopsies than in excisions in this study may reflect that pathologists were considering the larger clinical impact of an ADH diagnosis on a core biopsy (which would be followed by an excisional biopsy in the operating room) than on an excisional biopsy (no additional surgical management required) and more carefully considering the diagnosis. As physicians, pathologists should be able to use their clinical judgment when making diagnoses in grey zones (while using diagnostic criteria), to best serve patients in difficult clinical management decisions.

How should the pathologist approach intraductal proliferative lesions where the diagnosis of ADH is being considered? The cases from the B-Path study serve as well characterized examples of cases that highlight common issues in diagnostic agreement on these lesions. Based on these findings, a diagnostic approach that includes careful evaluation of the cytology, architectural patterns and extent of the lesion should be made in every case, with exclusion of other lesions in the differential when possible by ancillary studies. When features are borderline with another diagnosis in this spectrum (such as UDH or DCIS), this should be indicated in the report to communicate the borderline nature of the process to the clinical team and the patient. The clinical context should be considered and a conservative approach applied to core biopsy specimens. Second opinions can be sought to help ensure that diagnostic thresholds are uniform within a particular practice and with specialists in breast pathology. However, some cases will remain in the "grey zone," and the diagnosis of ADH should be understood by clinicians as a diagnosis with high inter-observer variability

for reasons that are not necessarily due to a higher frequency of “misdiagnoses,” but frequently due to the complexity and overlapping features preset in this spectrum of lesions.

Acknowledgments

This work was supported by the National Cancer Institute of the National Institutes of Health under award numbers R01 CA140560, R01 CA172343 and by the National Cancer Institute-funded Breast Cancer Surveillance Consortium award number HHSN261201100031C. The content is solely the responsibility of the authors and does not necessarily represent the views of the National Cancer Institute or the National Institutes of Health. The collection of cancer and vital status data used in this study was supported in part by several state public health departments and cancer registries throughout the U.S. For a full description of these sources, please see: <http://www.breastscreening.cancer.gov/work/acknowledgement.html>.

The authors wish to thank Ventana Medical Systems, Inc., a member of the Roche Group, for use of iScan Coreo Au™ whole slide imaging system, and HD View SL for the source code used to build our digital viewer. For a full description of HD View SL please see <http://hdviewsl.codeplex.com/>.^{a,b}

We would like to acknowledge all the pathologists who participated in the B-Path study.

References

1. Abdollahi A, Meysamie A, Sheikhabahaei S, et al. Inter/intra-observer reproducibility of Gleason scoring in prostate adenocarcinoma in Iranian pathologists. *Urology journal*. 2012; 9(2):486–90. [PubMed: 22641492]
2. Barnhill RL, Argenyi Z, Berwick M, et al. Atypical cellular blue nevi (cellular blue nevi with atypical features): lack of consensus for diagnosis and distinction from cellular blue nevi and malignant melanoma (“malignant blue nevus”). *The American journal of surgical pathology*. 2008; 32(1):36–44. [PubMed: 18162768]
3. Bean SM, Meara RS, Vollmer RT, et al. p16 Improves interobserver agreement in diagnosis of anal intraepithelial neoplasia. *Journal of lower genital tract disease*. 2009; 13(3):145–53. [PubMed: 19550211]
4. Coco DP, Goldblum JR, Hornick JL, et al. Interobserver variability in the diagnosis of crypt dysplasia in Barrett esophagus. *The American journal of surgical pathology*. 2011; 35(1):45–54. [PubMed: 21164286]
5. El-Zimaity HM, Wotherspoon A, de Jong D. Interobserver variation in the histopathological assessment of malt/malt lymphoma: towards a consensus. *Blood cells, molecules & diseases*. 2005; 34(1):6–16.
6. Fadare O, Parkash V, Dupont WD, et al. The diagnosis of endometrial carcinomas with clear cells by gynecologic pathologists: an assessment of interobserver variability and associated morphologic features. *The American journal of surgical pathology*. 2012; 36(8):1107–18. [PubMed: 22790851]
7. Fleskens SA, Bergshoeff VE, Voogd AC, et al. Interobserver variability of laryngeal mucosal premalignant lesions: a histopathological evaluation. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*. 2011; 24(7):892–8.
8. Foss FA, Milkins S, McGregor AH. Inter-observer variability in the histological assessment of colorectal polyps detected through the NHS Bowel Cancer Screening Programme. *Histopathology*. 2012; 61(1):47–52. [PubMed: 22486166]
9. Gerhard R, da Cunha Santos G. Inter- and intraobserver reproducibility of thyroid fine needle aspiration cytology: an analysis of discrepant cases. *Cytopathology : official journal of the British Society for Clinical Cytology*. 2007; 18(2):105–11. [PubMed: 17397495]
10. Gupta T, Nair V, Epari S, Pietsch T, Jalali R. Concordance between local, institutional, and central pathology review in glioblastoma: implications for research and practice: a pilot study. *Neurology India*. 2012; 60(1):61–5. [PubMed: 22406782]
11. Kerkhof M, van Dekken H, Steyerberg EW, et al. Grading of dysplasia in Barrett’s oesophagus: substantial interobserver variation between general and gastrointestinal pathologists. *Histopathology*. 2007; 50(7):920–7. [PubMed: 17543082]

12. Longnecker DS, Adsay NV, Fernandez-del Castillo C, et al. Histopathological diagnosis of pancreatic intraepithelial neoplasia and intraductal papillary-mucinous neoplasms: interobserver agreement. *Pancreas*. 2005; 31(4):344–9. [PubMed: 16258368]
13. Parkash V, Bifulco C, Feinn R, Concato J, Jain D. To count and how to count, that is the question: interobserver and intraobserver variability among pathologists in lymph node counting. *American journal of clinical pathology*. 2010; 134(1):42–9. [PubMed: 20551265]
14. Puppa G, Senore C, Sheahan K, et al. Diagnostic reproducibility of tumour budding in colorectal cancer: a multicentre, multinational study using virtual microscopy. *Histopathology*. 2012
15. Stang A, Trocchi P, Ruschke K, et al. Factors influencing the agreement on histopathological assessments of breast biopsies among pathologists. *Histopathology*. 2011; 59(5):939–49. [PubMed: 22092405]
16. Anderson TJ, Sufi F, Ellis IO, Sloane JP, Moss S. Implications of pathologist concordance for breast cancer assessments in mammography screening from age 40 years. *Human pathology*. 2002; 33(3):365–71. [PubMed: 11979379]
17. Beck JS. Observer variability in reporting of breast lesions. *Journal of clinical pathology*. 1985; 38(12):1358–65. [PubMed: 3001153]
18. Bianchi S, Palli D, Galli M, et al. Reproducibility of histological diagnoses and diagnostic accuracy of non palpable breast lesions. *Pathology, research and practice*. 1994; 190(1):69–76.
19. Bodian CA, Perzin KH, Lattes R, Hoffmann P. Reproducibility and validity of pathologic classifications of benign breast disease and implications for clinical applications. *Cancer*. 1993; 71(12):3908–13. [PubMed: 8508356]
20. Collins LC, Connolly JL, Page DL, et al. Diagnostic agreement in the evaluation of image-guided breast core needle biopsies: results from a randomized clinical trial. *The American journal of surgical pathology*. 2004; 28(1):126–31. [PubMed: 14707874]
21. Elston CW, Sloane JP, Amendoeira I, et al. Causes of inconsistency in diagnosing and classifying intraductal proliferations of the breast. *European Commission Working Group on Breast Screening Pathology*. *European Journal of Cancer*. 2000; 36(14):1769–72. [PubMed: 10974624]
22. Ghofrani M, Tapia B, Tavassoli FA. Discrepancies in the diagnosis of intraductal proliferative lesions of the breast and its management implications: results of a multinational survey. *Virchows Archiv : an international journal of pathology*. 2006; 449(6):609–16. [PubMed: 17058097]
23. Palazzo JP, Hyslop T. Hyperplastic ductal and lobular lesions and carcinomas in situ of the breast: reproducibility of current diagnostic criteria among community- and academic-based pathologists. *The breast journal*. 1998; 4(4):230–7. [PubMed: 21223441]
24. Palli D, Galli M, Bianchi S, et al. Reproducibility of histological diagnosis of breast lesions: results of a panel in Italy. *European Journal of Cancer*. 1996; 32A(4):603–7. [PubMed: 8695260]
25. Rosai J. Borderline epithelial lesions of the breast. *The American journal of surgical pathology*. 1991; 15(3):209–21. [PubMed: 1847606]
26. Schnitt SJ, Connolly JL, Tavassoli FA, et al. Interobserver reproducibility in the diagnosis of ductal proliferative breast lesions using standardized criteria. *The American journal of surgical pathology*. 1992; 16(12):1133–43. [PubMed: 1463092]
27. Sloane JP, Ellman R, Anderson TJ, et al. Consistency of histopathological reporting of breast lesions detected by screening: findings of the U.K. National External Quality Assessment (EQA) Scheme. U. K. National Coordinating Group for Breast Screening Pathology. *European Journal of Cancer*. 1994; 30A(10):1414–9. [PubMed: 7833094]
28. Wells WA, Carney PA, Eliassen MS, Tosteson AN, Greenberg ER. Statewide study of diagnostic agreement in breast pathology. *Journal of the National Cancer Institute*. 1998; 90(2):142–5. [PubMed: 9450574]
29. Jain RK, Mehta R, Dimitrov R, et al. Atypical ductal hyperplasia: interobserver and intraobserver variability. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*. 2011; 24(7):917–23.
30. Perkins C, Balma D, Garcia R. Why current breast pathology practices must be evaluated. A Susan G. Komen for the Cure white paper: June 2006. *The breast journal*. 2007; 13(5):443–7. [PubMed: 17760664]
31. Saul S. Prone to error: Earliest steps to find cancer. *New York Times A*. 2010:1.

32. Tanner L. Breast tissue misdiagnosed more often than you think. *The Globe and Mail*. 2015 Mar 18.2015
33. Allison KH, Reisch LM, Carney PA, et al. Understanding diagnostic variability in breast pathology: lessons learned from an expert consensus review panel. *Histopathology*. 2014; 65(2): 240–51. [PubMed: 24511905]
34. Elmore JG, Longton GM, Carney PA, et al. Diagnostic concordance among pathologists interpreting breast biopsy specimens. *Jama*. 2015; 313(11):1122–32. [PubMed: 25781441]
35. Oster NV, Carney PA, Allison KH, et al. Development of a diagnostic test set to assess agreement in breast pathology: practical application of the Guidelines for Reporting Reliability and Agreement Studies (GRRAS). *BMC women's health*. 2013; 13:3. [PubMed: 23379630]
36. Nelson HD, Weerasinghe R, Martel M, et al. Development of an electronic breast pathology database in a community health system. *Journal of pathology informatics*. 2014; 5:26. [PubMed: 25191625]
37. Geller BM, Nelson HD, Carney PA, et al. Second opinion in breast pathology: policy, practice and perception. *J Clin Pathol*. 2014; 67(11):955–60. [PubMed: 25053542]
38. Jara-Lazaro AR, Tan PH. Pattern and spectrum of morphology referrals in breast pathology consultation. *Pathology*. 2008; 40(6):564–72. [PubMed: 18752122]
39. Putti TC, Pinder SE, Elston CW, Lee AH, Ellis IO. Breast pathology practice: most common problems in a consultation service. *Histopathology*. 2005; 47(5):445–57. [PubMed: 16241992]
40. Lacroix-Triki M, Mery E, Voigt JJ, Istier L, Rochaix P. Value of cytokeratin 5/6 immunostaining using D5/16 B4 antibody in the spectrum of proliferative intraepithelial lesions of the breast. A comparative study with 34betaE12 antibody. *Virchows Arch*. 2003; 442(6):548–54. [PubMed: 12712335]
41. Otterbach F, Bankfalvi A, Bergner S, Decker T, Krech R, Boecker W. Cytokeratin 5/6 immunohistochemistry assists the differential diagnosis of atypical proliferations of the breast. *Histopathology*. 2000; 37(3):232–40. [PubMed: 10971699]
42. Shoker BS, Jarvis C, Sibson DR, Walker C, Sloane JP. Oestrogen receptor expression in the normal and pre-cancerous breast. *The Journal of pathology*. 1999; 188(3):237–44. [PubMed: 10419589]
43. Jain RK, Mehta R, Dimitrov R, et al. Atypical ductal hyperplasia: interobserver and intraobserver variability. *Mod Pathol*. 2011; 24(7):917–23. [PubMed: 21532546]
44. Dabbs DJ, Schnitt SJ, Geyer FC, et al. Lobular neoplasia of the breast revisited with emphasis on the role of E-cadherin immunohistochemistry. *Am J Surg Pathol*. 2013; 37(7):e1–11. [PubMed: 23759937]

Specimen type: Core Excision

SCREENING:

Number of regions of potential interest to screen: 1-2 3-5 >5

Diagnostic area obvious on low power: Yes No Borderline

Distracting diagnoses present in other areas? No Yes: Lobular lesion(s) FEA
 UDH Papilloma Other: _____

Are there other diagnoses in the differential or present *within* the lesion of interest?: No

Yes: Lobular process FEA Intraductal papilloma UDH

Other: _____

EXTENT/LOW POWER FEATURES:

Is the lesion of interest in a papillary lesion? Yes No

Extent of lesion(s) of interest (ADH or IPA):

Number of separate areas with atypia/lesion containing atypia:

Largest single area with atypia/lesions containing atypia:

Number of foci (basement membrane bound spaces) with atypia diagnosis: 1-2 3-5 6+

Largest single discrete focus: ≤ 2mm > 2 mm

HIGHER POWER FEATURES:

Cytology of lesion of interest:

Monotony: Very monotonous Questionably monotonous: features borderline with UDH
 features borderline with intermediate grade DCIS

Apocrine cytology: Present Not present

Nuclei hyperchromatic: Present Not present

Nucleoli prominent: Present Not present

Calcifications associated with lesion of interest: Yes No

Architecture of lesion of interest:

Predominant type of architecture present: Cribriform Micropapillary Solid

Papillary Flat pattern

Development of architecture: Well-developed atypical architecture (polarized cribriform spaces/microacini, club-shaped structures, etc) Partially developed atypical architecture Solid lesion No/subtle architectural atypia

Uniformity of cytologic and architectural features throughout the lesion: Lesion uniformly involved by process Lesion contains mixed areas with and without atypical features

Figure 1.
Scoring Form for Histologic Features Evaluated in 72 Expert Consensus ADH Cases

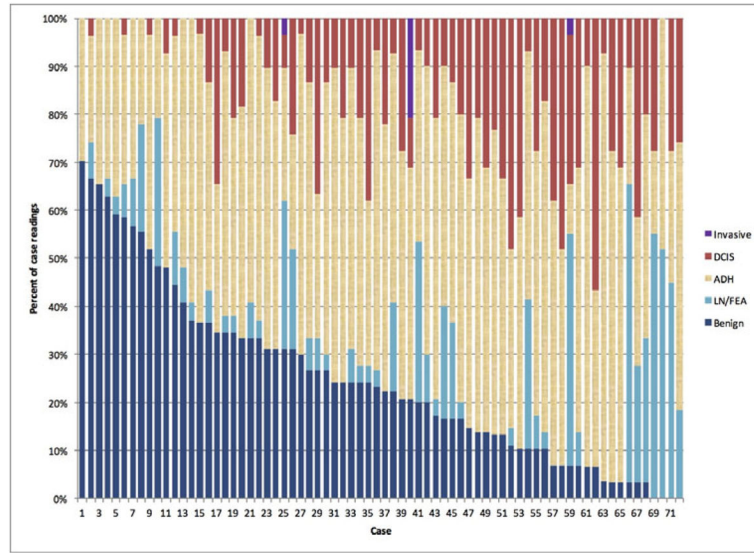


Figure 2. Spectrum of Diagnoses Recorded by Participating Pathologists in the 72 Cases with an Expert Consensus Diagnosis of ADH

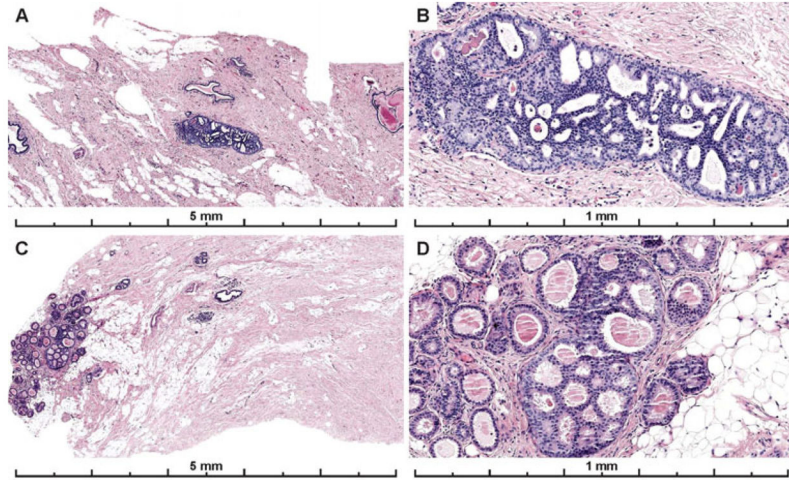


Figure 3. Photos of the diagnostic areas from the two cases with the highest participant agreement with the consensus diagnosis of ADH. Both cases had focal lesions, less than 2 mm (and < 2 membrane bound spaces), that were obvious on low power, had obvious cytologic monotony and a cribriform architectural pattern. A–B) A case with 83% of participant agreement with the expert consensus diagnosis of ADH; 7% recorded a Benign diagnosis (UDH), and 10% recorded a DCIS diagnosis for the case (30 total interpretations). C–D) 89% agreed with the consensus diagnosis of ADH; 4% recorded a Benign diagnosis (UDH) and 7% recorded a DCIS diagnosis on the case (27 total interpretations). Lesions with these features should be reproducibly classified as ADH as serve as good examples of this diagnosis.

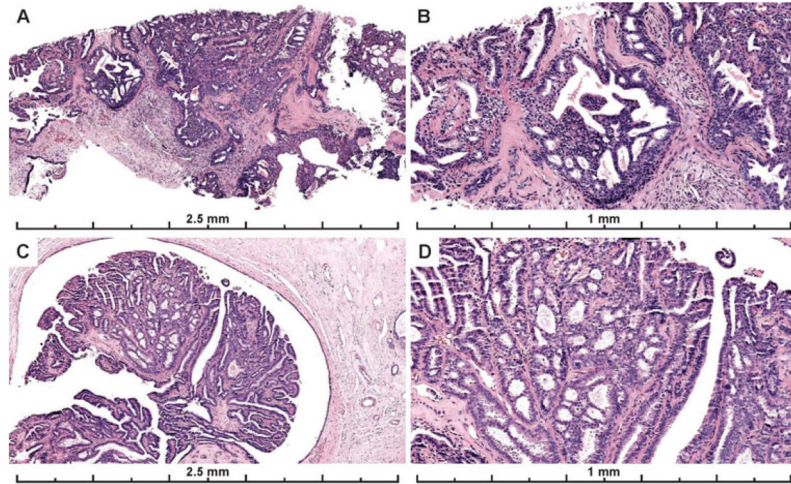


Figure 4. Papillary lesions with ADH were associated with higher diagnostic agreement than ADH not in a papillary lesion. A–B) 67% of participants recorded a diagnosis of ADH for this papillary lesion; 28% recorded a DCIS diagnosis and 3% a Benign diagnosis (mostly papilloma without atypia) (29 interpretations). C–D) 59% of participants recorded a diagnosis of ADH for this papillary lesion; 31% recorded a Benign diagnosis (mostly papilloma without atypia), and 10% recorded a DCIS diagnosis for the case (29 total interpretations). The first case (A–B) had less discrete areas of atypia in more than one area of the papilloma, making it borderline with DCIS. However, each focus was considered < 2mm and the consensus diagnosis was ADH on this core needle biopsy sample. The second case has a single, more discrete area of cribriform architecture with cytologic monotony within the papillary proliferation measuring < 2 mm. These areas appear distinct from the background normal papillary epithelium or UDH but due to their limited in extent (< 2 mm), they fall short of most pathologists’ threshold for DCIS.

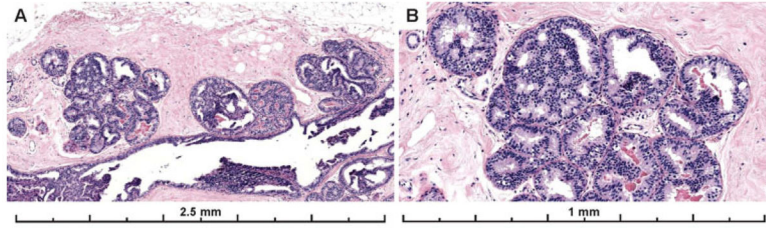


Figure 5. Example of a case with obvious cytologic monotony and cribriform architecture that was close to the 2 mm threshold distinguishing between a diagnosis of ADH versus limited extent low grade DCIS. 60% of participants recorded a diagnosis of ADH, 20% recorded a DCIS, 17% recorded a Benign diagnosis, and 3% an FEA diagnosis for the case (30 total interpretations). When a lesion with these features is close to the extent threshold, diagnostic disagreement is predictable. The lesion involves more than two membrane bound spaces but is close to 2 mm in extent. The expert consensus panel (and majority of participants) applied a more conservative approach with this borderline lesion, classifying it as ADH rather than low grade DCIS. A comment can be included in the report that the lesion is borderline with low grade DCIS.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

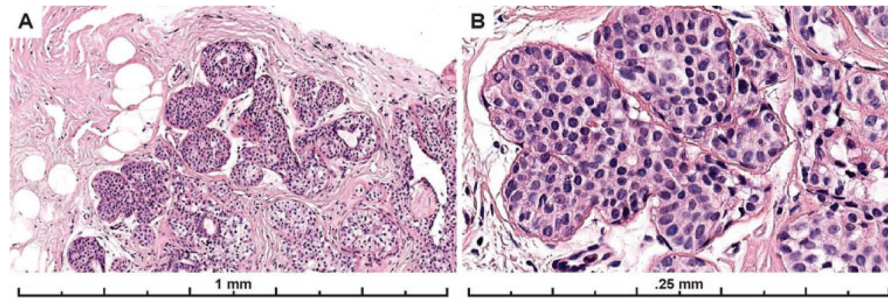


Figure 6. Example images of the diagnostic area from a case with solid architecture and subtle cytologic monotony. Only 17% of participating pathologists recorded a diagnosis of ADH, with 55% recording the case as LN and 28% as DCIS (29 total interpretations). While the differential diagnosis includes a lobular in situ lesion (ALH/LCIS) that may be resolved with an E-cadherin stain, subtle micro-acini supporting a ductal process are evident on the H&E at high power (panel B).

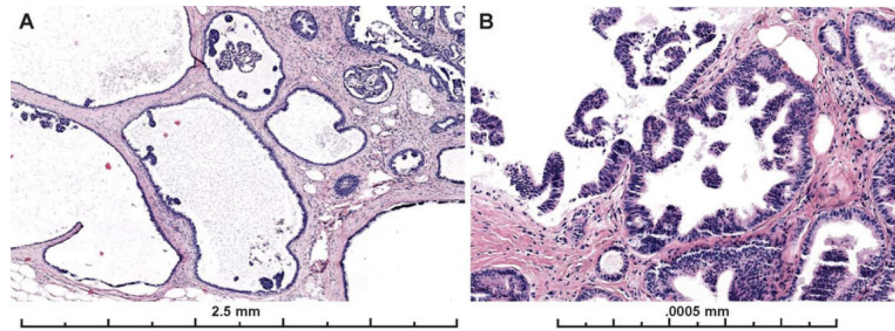


Figure 7.

Examples images from two cases with micropapillary architecture, which was associated with worse diagnostic agreement with the expert consensus diagnosis of ADH. A) Only 28% of participants recorded a diagnosis of ADH on this case, 45% recorded a LN (there was LN also present in multiple foci elsewhere on the slide), 28% recorded a DCIS diagnosis and 0% recorded a Benign diagnosis (29 interpretations). B) Only 24% of participants recorded this case as ADH, with 31% as Benign (mostly columnar cell hyperplasia and UDH), 21% recording the case as FEA and 24% as DCIS (29 total interpretations). Both lesions have scattered micropapillary, club-shaped structures that are only partially involving dilated spaces involved by FEA. Micropapillary patterns can be subtle and easily missed if focal. In the first case (A), the concurrent diagnosis of lobular neoplasia in multiple areas of the same slide may have been a distractor.

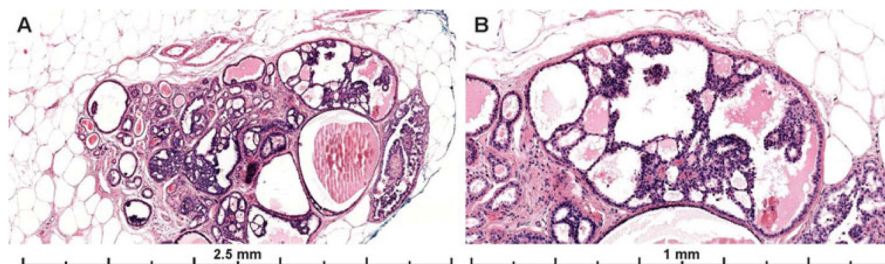


Figure 8.

Example images from a case with a single region of potential interest to screen but borderline cytologic monotony (ADH vs UDH) and architectural changes that appear cribriform but lack polarized spaces. 31% of participants recorded an ADH diagnosis for the case, 59% recorded a Benign diagnosis (94% UDH) and 7% recorded a LN diagnosis, and 3% recorded a DCIS diagnosis (29 total interpretations). When the both the architectural and the cytologic features are borderline between UDH and ADH, diagnostic disagreement is more likely. Additional levels, immunohistochemistry (CK5/6, ER or the ADH5 stain) and additional opinions can be helpful in this differential if these findings are present on a core needle biopsy where it would be most clinically relevant. Given the findings on the single H&E slide available in this test set, the expert panel classified this lesion as ADH based on the architectural atypia present and subtle monotony. However, the findings are borderline with UDH, which can be mentioned in the report.

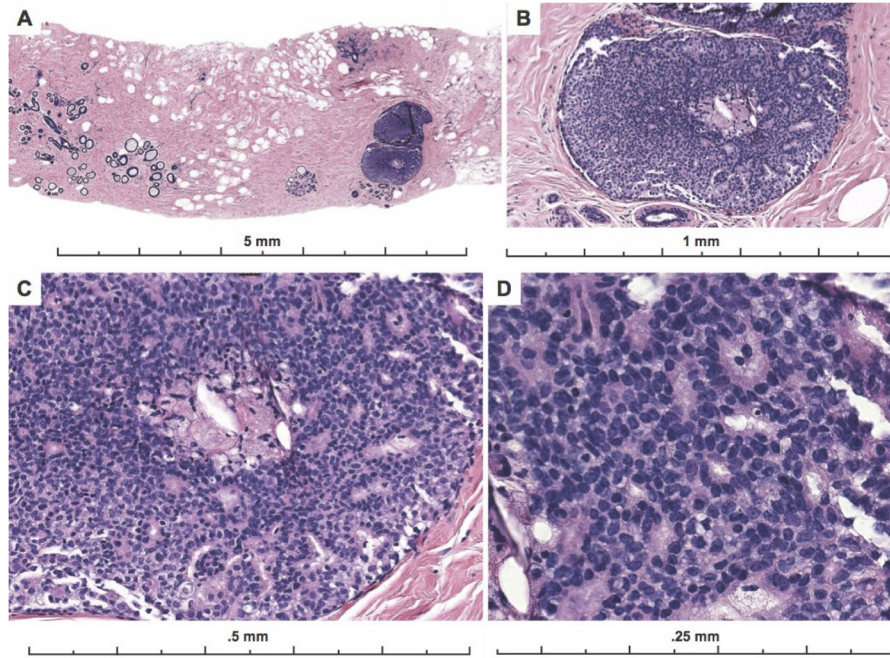


Figure 9. Example images from a case with a single region of potential interest to screen but borderline monotony (UDH vs ADH) and solid/subtle architecture. On higher power (panels C and D) subtle microacini are apparent with polarization of cells towards the lumen. This case had an almost even distribution between three diagnostic categories with 31% of participants recording an ADH diagnosis, 34% recording a Benign diagnosis (100% recorded UDH), and 34% recording a DCIS for the case (29 total interpretations). The differential in this case includes UDH due to the nuclear crowding and slit like spaces but also ADH because of the subtle polarized spaces being formed. The subtly hyperchromatic nuclear cytology also raises the differential of a low-intermediate grade DCIS. Additional levels, immunohistochemistry (CK5/6, ER or the ADH5 stain) and additional opinions can be helpful in this differential. Based on the presence of a single lesion measuring < 2 mm and involving only two membrane bound spaces on this H&E alone, classification as ADH was considered the best diagnosis by the expert consensus panel.

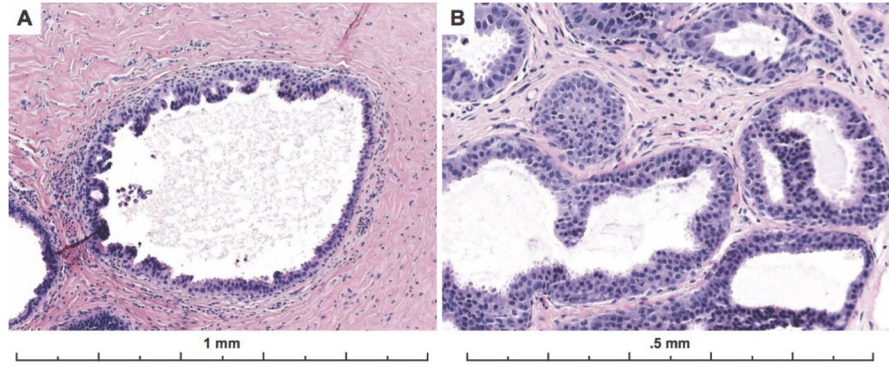


Figure 10. Example images from two cases that had a predominantly flat pattern of architecture with only subtle early arches, bridges or micropapillations that the expert consensus panel classified as ADH. A) The cytology is monotonous, with rounded nuclei, and a predominantly flat growth pattern. 33% of participants recorded a highest order diagnosis of FEA in this case, 40% as ADH, 20% as benign, and 7% as DCIS. B) Similarly, the cytology of the second case is monotonous with predominantly a flat growth pattern but very focal, early micropapillae and bridges being formed, consistent with ADH. 23% of participants recorded a highest order diagnosis of FEA in this case, 50% as ADH, 17% as Benign, and 10% as DCIS. When FEA is present, the case should be closely examined for subtle architectural patterns consistent with an ADH diagnosis.

Participant ratings of diagnostic challenges present in consensus atypia cases and association with agreement with reference diagnosis

Table 1

Participant Rating	Total	Agree with reference diagnosis of ADH ^b N (column %)	P-value ^d	Called DCIS ^b N %	Called FEA or LN ^b N %	Called Benign N %	Called Invasive N %
All interpretations	2070	990 (48%)		353 (17%)	207 (10%)	512 (25%)	8 (0%)
Difficulty rating							
Low difficulty (Likert 1–3)	1073 (52%)	426 (43%) (40%)	<.001	153 (43%) (14%)	115 (56%) (11%)	374 (73%) (35%)	5 (63%) (0%)
High difficulty (Likert 4–6)	997 (48%)	564 (57%) (57%)		200 (57%) (20%)	92 (44%) (9%)	138 (27%) (14%)	3 (38%) (0%)
Borderline with another diagnosis							
Yes	937 (45%)	545 (55%) (58%)	<.001	180 (51%) (19%)	71 (34%) (8%)	138 (27%) (15%)	3 (38%) (0%)
No	1133 (55%)	445 (45%) (39%)		173 (49%) (15%)	136 (66%) (12%)	374 (73%) (33%)	5 (63%) (0%)
Second opinion desired (or policy)							
Yes	1664 (80%)	864 (87%) (52%)	<.001	320 (91%) (19%)	177 (86%) (11%)	295 (58%) (18%)	8 (100%) (0%)
No	406 (20%)	126 (13%) (31%)		33 (9%) (8%)	30 (14%) (7%)	217 (42%) (53%)	0 (0%) (0%)
Confidence rating							
High confidence (Likert 1–3)	1483 (72%)	668 (67%) (45%)	<.001	236 (67%) (16%)	153 (74%) (10%)	421 (82%) (28%)	5 (63%) (0%)
Low confidence (Likert 4–6)	587 (28%)	322 (33%) (55%)	117 (33%) (20%)	54 (26%) (9%)	91 (18%) (16%)	3 (38%) (1%)	

^a Multinomial regression from generalized estimating equations (GEE) model accounting for repeated measures with binomial distribution for the dependent variable. This P-value compares diagnosis of ADH to diagnosis of Benign

^b ADH=Atypical ductal hyperplasia and intraductal papilloma with atypical ductal hyperplasia; FEA= Flat epithelial atypia; LN=Lobular neoplasia; DCIS= Ductal carcinoma in situ

^c Note: date 18may16 - sas release: 9.4 - a023_table1_revise_again.sas - sue peacock (206) 744-9912 peacocks@uw.edu GIM

Table 2
Association of Histologic Features Present with Participant Agreement with Expert Consensus Diagnosis of ADH in 72 Cases

Lesion Features	Number of cases	Number of assessments	Participant diagnosis of ADH in agreement with expert consensus diagnosis of ADH?		OR (CI)	P-value ^a
			Agree N %	Disagree N %		
Total	72	2070 (100%)	990 (48%)	1080 (52%)		
Papillary lesion?						<.001
No	58	1667 (100%)	765 (46%)	902 (54%)	1.0	
Yes	14	403 (100%)	225 (56%)	178 (44%)	1.49 (1.19, 1.87)	
Architecture pattern present in lesion:						<.001
Cribriform	55	1579 (100%)	801 (51%)	778 (49%)	1.0	
Flat	3	89 (100%)	47 (53%)	42 (47%)	1.09 (0.69, 1.71)	
Micro-papillary	9	259 (100%)	103 (40%)	156 (60%)	0.64 (0.50, 0.83)	
Solid	5	143 (100%)	39 (27%)	104 (73%)	0.36 (0.25, 0.54)	
Specimen Type:						0.009
Core	48	1381 (100%)	685 (50%)	696 (50%)	1.0	
Excision	24	689 (100%)	305 (44%)	384 (56%)	0.81 (0.69, 0.95)	
Cytologic monotony in lesion:						0.013
Very monotonous	52	1495 (100%)	742 (50%)	753 (50%)	1.0	
Not monotonous or Borderline	20	575 (100%)	248 (43%)	327 (57%)	0.77 (0.63, 0.95)	
Diagnostic area obvious on low power?^b:						0.020
No	15	431 (100%)	193 (45%)	238 (55%)	1.0	
Yes	48	1381 (100%)	687 (50%)	694 (50%)	1.22 (1.00, 1.49)	
Borderline	8	229 (100%)	96 (42%)	133 (58%)	0.89 (0.65, 1.22)	
Number of foci with atypia:						0.041
1–2 foci	21	601 (100%)	264 (44%)	337 (56%)	1.0	
3–5 foci	29	834 (100%)	420 (50%)	414 (50%)	1.30 (1.06, 1.58)	
6+ foci	22	635 (100%)	306 (48%)	329 (52%)	1.19 (0.94, 1.49)	
Number regions of interest (ROI) to screen^b:						0.26
1–2 foci	30	859 (100%)	420 (49%)	439 (51%)	1.0	

Lesion Features	Number of cases	Number of assessments	Participant diagnosis of ADH in agreement with expert consensus diagnosis of ADH?		OR (CI)	P-value ^d
			Agree N %	Disagree N %		
3–5 foci	30	869 (100%)	417 (48%)	452 (52%)	0.96 (0.82, 1.14)	
>5 foci	11	313 (100%)	138 (44%)	175 (56%)	0.82 (0.65, 1.04)	
Distracting Areas?						0.19
No	53	1528 (100%)	743 (49%)	785 (51%)	1.0	
Yes	19	542 (100%)	247 (46%)	295 (54%)	0.88 (0.74, 1.06)	
Other diagnosis in Ddx or within the lesion^b?						0.20
No	23	662 (100%)	329 (50%)	333 (50%)	1.0	
Yes	48	1379 (100%)	647 (47%)	732 (53%)	0.89 (0.75, 1.06)	
Number of areas with atypia^b:						0.52
1	46	1324 (100%)	620 (47%)	704 (53%)	1.0	
2	17	493 (100%)	246 (50%)	247 (50%)	1.13 (0.90, 1.41)	
3+	7	199 (100%)	99 (50%)	100 (50%)	1.12 (0.82, 1.55)	
Largest area with atypia^b:						0.26
<2.5	34	979 (100%)	460 (47%)	519 (53%)	1.0	
2.5	37	1062 (100%)	524 (49%)	538 (51%)	1.10 (0.93, 1.29)	
Largest single discrete focus:						0.52
2mm	58	1665 (100%)	791 (48%)	874 (52%)	1.0	
> 2mm	14	405 (100%)	199 (49%)	206 (51%)	1.07 (0.87, 1.30)	
Apocrine cytology?						0.14
No	67	1927 (100%)	914 (47%)	1013 (53%)	1.0	
Yes	5	143 (100%)	76 (53%)	67 (47%)	1.26 (0.93, 1.70)	
Hyperchromatic nuclei?						0.19
No	57	1631 (100%)	793 (49%)	838 (51%)	1.0	
Yes	15	439 (100%)	197 (45%)	242 (55%)	0.86 (0.69, 1.08)	
Nucleoli present?						0.84
No	59	1695 (100%)	809 (48%)	886 (52%)	1.0	
Yes	13	375 (100%)	181 (48%)	194 (52%)	1.02 (0.83, 1.25)	

Lesion Features	Number of cases	Number of assessments	Participant diagnosis of ADH in agreement with expert consensus diagnosis of ADH?		OR (CI)	P-value ^d
			Agree N %	Disagree N %		
Calcifications present in lesion?						
No	43	1235 (100%)	601 (49%)	634 (51%)	1.0	0.35
Yes	29	835 (100%)	389 (47%)	446 (53%)	0.92 (0.77, 1.10)	
Development of architecture in lesion:						0.59
Well developed	14	409 (100%)	189 (46%)	220 (54%)	1.0	
Partial or Solid/Subtle	56	1604 (100%)	768 (48%)	836 (52%)	1.07 (0.84, 1.37)	
Uniformity of process in lesion:						0.79
Uniform	53	1519 (100%)	729 (48%)	790 (52%)	1.0	
Partial	18	522 (100%)	247 (47%)	275 (53%)	0.97 (0.80, 1.19)	

^aOdds ratios and p-values produced using GEE.

^bFor these features, there were rare cases in which there was no agreement on the scoring (using the 3 reviewing pathologist scores). This occurred in < 1% of all assessments (8/1224) and they were not included in the overall scoring for that histologic feature.

Table 3

Histologic Features Present in Cases Identified as ADH by Consensus Experts (N=72) Stratified by Participant Diagnosis

Lesion Features	Number of cases	Number of assessments N (Row %)	Study Participant Diagnosis				
			Benign	FEA or LN	ADH	DCIS	Invasive
Total	72	2070 (100%)	512 (25%)	207 (10%)	990 (48%)	353 (17%)	8 (<0.1%)
Papillary lesion?							
No	58	1667 (100%)	421 (25%)	204 (12%)	765 (46%)	275 (16%)	2 (<0.1%)
Yes	14	403 (100%)	91 (23%)	3 (1%)	225 (56%)	78 (19%)	6 (1%)
Architecture pattern present in lesion:							
Cribiform	55	1579 (100%)	429 (27%)	107 (7%)	801 (51%)	235 (15%)	7 (<0.1%)
Flat	3	89 (100%)	16 (18%)	13 (15%)	47 (53%)	13 (15%)	0 (<0.1%)
Micropapillary	9	259 (100%)	54 (21%)	34 (13%)	103 (40%)	68 (26%)	0 (<0.1%)
Solid	5	143 (100%)	13 (9%)	53 (37%)	39 (27%)	37 (26%)	1 (1%)
Specimen Type:							
Core	48	1381 (100%)	328 (24%)	108 (8%)	685 (50%)	253 (18%)	7 (1%)
Excision	24	689 (100%)	184 (27%)	99 (14%)	305 (44%)	100 (15%)	1 (<0.1%)
Cytologic monotony in lesion:							
Very	52	1495 (100%)	322 (22%)	152 (10%)	742 (50%)	272 (18%)	7 (<0.1%)
Borderline with UDH	20	575 (100%)	190 (33%)	55 (10%)	248 (43%)	81 (14%)	1 (<0.1%)
Diagnostic area obvious on low power?^a							
No	15	431 (100%)	130 (30%)	61 (14%)	193 (45%)	46 (11%)	1 (<0.1%)
Yes	48	1381 (100%)	321 (23%)	88 (6%)	687 (50%)	279 (20%)	6 (<0.1%)
Borderline	8	229 (100%)	61 (27%)	43 (19%)	96 (42%)	28 (12%)	1 (<0.1%)
Number of foci with atypia:							
1–2 foci	21	601 (100%)	215 (36%)	76 (13%)	264 (44%)	45 (7%)	1 (<0.1%)
3–5 foci	29	834 (100%)	213 (26%)	67 (8%)	420 (50%)	133 (16%)	1 (<0.1%)
6+ foci	22	635 (100%)	84 (13%)	64 (10%)	306 (48%)	175 (28%)	6 (1%)
Number regions of interest (ROI) to screen^a:							
1–2 foci	30	859 (100%)	222 (26%)	85 (10%)	420 (49%)	126 (15%)	6 (1%)

Lesion Features	Number of cases	Number of assessments N (Row %)	Study Participant Diagnosis				
			Benign	FEA or LN	ADH	DCIS	Invasive
3-5 foci	30	869 (100%)	196 (23%)	79 (9%)	417 (48%)	175 (20%)	2 (<0.1%)
>5 foci	11	313 (100%)	91 (29%)	34 (11%)	138 (44%)	50 (16%)	0 (<0.1%)
Distracting Areas?							
No	53	1528 (100%)	379 (25%)	137 (9%)	743 (49%)	267 (17%)	2 (<0.1%)
Yes	19	542 (100%)	133 (25%)	70 (13%)	247 (46%)	86 (16%)	6 (1%)
Other diagnosis in Ddx or within the lesion ²:							
No	23	662 (100%)	198 (30%)	34 (5%)	329 (50%)	101 (15%)	0 (<0.1%)
Yes	48	1379 (100%)	308 (22%)	173 (13%)	647 (47%)	249 (18%)	2 (<0.1%)
Number of areas with atypia ²:							
1	46	1324 (100%)	343 (26%)	164 (12%)	620 (47%)	196 (15%)	1 (<0.1%)
2	17	493 (100%)	113 (23%)	36 (7%)	246 (50%)	91 (18%)	7 (1%)
3+	7	199 (100%)	34 (17%)	7 (4%)	99 (50%)	59 (30%)	0 (<0.1%)
Largest area with atypia ²:							
<2.5	34	979 (100%)	264 (27%)	119 (12%)	460 (47%)	134 (14%)	2 (<0.1%)
2.5	37	1062 (100%)	234 (22%)	79 (7%)	524 (49%)	219 (21%)	6 (1%)
Largest single discrete focus:							
2mm	58	1665 (100%)	433 (26%)	180 (11%)	791 (48%)	259 (16%)	2 (<0.1%)
> 2mm	14	405 (100%)	79 (20%)	27 (7%)	199 (49%)	94 (23%)	6 (1%)
Apocrine cytology?							
No	67	1927 (100%)	485 (25%)	198 (10%)	914 (47%)	322 (17%)	8 (<0.1%)
Yes	5	143 (100%)	27 (19%)	9 (6%)	76 (53%)	31 (22%)	0 (<0.1%)
Hyperchromatic nuclei?							
No	57	1631 (100%)	428 (26%)	158 (10%)	793 (49%)	245 (15%)	7 (<0.1%)
Yes	15	439 (100%)	84 (19%)	49 (11%)	197 (45%)	108 (25%)	1 (<0.1%)
Nucleoli present?							
No	59	1695 (100%)	436 (26%)	162 (10%)	809 (48%)	287 (17%)	1 (<0.1%)
Yes	13	375 (100%)	76 (20%)	45 (12%)	181 (48%)	66 (18%)	7 (2%)
Calcifications present in lesion?							

Lesion Features	Number of cases	Number of assessments N (Row %)	Study Participant Diagnosis				
			Benign	FEA or LN	ADH	DCIS	Invasive
No	43	1235 (100%)	345 (28%)	106 (9%)	601 (49%)	175 (14%)	8 (1%)
Yes	29	835 (100%)	167 (20%)	101 (12%)	389 (47%)	178 (21%)	0 (<0.1%)
Development of architecture in lesion:							
Well/Solid	14	409 (100%)	59 (14%)	51 (12%)	189 (46%)	103 (25%)	7 (2%)
Partial/No or Subtle	56	1604 (100%)	443 (28%)	151 (9%)	768 (48%)	241 (15%)	1 (<0.1%)
Uniformity of process in lesion:							
Uniform	53	1519 (100%)	393 (26%)	148 (10%)	729 (48%)	242 (16%)	7 (<0.1%)
Partial	18	522 (100%)	119 (23%)	44 (8%)	247 (47%)	111 (21%)	1 (<0.1%)

FEA = flat epithelial atypia, LN = lobular neoplasia (atypical lobular hyperplasia or lobular carcinoma in situ), ADH= atypical ductal hyperplasia and intraductal papilloma with atypical ductal hyperplasia, DCIS = ductal carcinoma in situ.

⁴For these features, there were rare cases in which there was no agreement on the scoring (using the 3 reviewing pathologist scores). This occurred in < 1% of all assessments (8/1224) and they were not included in the overall scoring for that histologic feature.