Interobserver reproducibility of cervical histology interpretation with and without p16 immunohistochemistry

Amy S. Tao, BA,[1] Rosemary Zuna, MD,[2] Teresa M. Darragh, MD,[3] Niels Grabe, PhD,[4] Bernd Lahrmann, PhD,[4] Megan A. Clarke, PhD,[1] Nicolas Wentzensen, MD, PhD, MS[1]

[1]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services (DHHS), Bethesda, MD, US; [2]Department of Pathology, University of Oklahoma Health Sciences Center, Oklahoma City, OK, US; [3]Department of Pathology, University of California San Francisco, San Francisco, CA, US; [4]Department of Pathology, University Medical Center Göttingen, Göttingen, Germany

ABSTRACT

Objectives: Histopathological diagnosis of colposcopically identified cervical lesions is a critical step for the recognition of cervical cancer precursors requiring treatment. Although there have been efforts to standardize the histologic diagnosis of cervical biopsy specimens, in terms of terminology and use of biomarkers, there is no uniform approach in the pathology community. Adjunctive p16 immunohistochemistry (IHC) can highlight precancer diagnoses, with use recommendations outlined by the Lower Anogenital Squamous Terminology project.

Methods: We assessed the diagnostic reproducibility of cervical histopathological biopsy specimens with and without p16 staining among 2 expert pathologists.

Results: Interpretation of p16 IHC as positive vs negative was highly reproducible (92.5% agreement, $\kappa = 0.85$); greater variation was seen in the choice of which biopsy specimens required adjunctive p16 staining (78.0% agreement, $\kappa = 0.43$). Adjunctive p16 IHC did not significantly increase diagnostic agreement under multitiered grading systems (benign vs cervical intraepithelial neoplasia [CIN] 1/low-grade squamous intraepithelial lesion vs atypical squamous metaplasia vs CIN2/high-grade squamous intraepithelial lesion [HSIL] vs CIN3/HSIL-CIN3 vs cancer) (65.5% agreement, $\kappa = 0.56$ without p16; 70.0% agreement, $\kappa = 0.58$ with p16). However, when dichotomizing diagnoses based on clinical management (less than HSIL vs HSIL+), diagnostic agreement increased with p16 IHC (90.5% agreement, $\kappa = 0.79$ without p16; 92.0% agreement, $\kappa = 0.84$ with p16). For biopsy specimens taken from women positive for human papillomavirus (HPV) type 16, agreement was similar with or without adjunctive p16 ($\kappa = 0.80$ without p16; $\kappa = 0.78$-0.80 with p16). In contrast, p16 IHC substantially improved diagnostic agreement for cervical biopsy specimens taken from women positive for other high-risk HPV strains, producing improvements in κ from 0.03 to 0.24.

Conclusions: Adjunctive p16 immunostaining provides useful information in the evaluation of cervical biopsies for precancer. In our study, we have demonstrated that it is highly reproducible between 2 pathologists, although the decision of which biopsies warrant its use is less so. Furthermore, although p16 IHC showed a limited increase in diagnostic reproducibility for all biopsies included in our study, it did demonstrate a more sizable gain in biopsies negative for HPV 16 but positive for other high-risk genotypes. Further studies are needed to clarify the role of p16 IHC and how it can be optimized for the detection of

KEY POINTS

- Interpretation of p16 immunohistochemistry (IHC) as positive vs negative was highly reproducible between 2 pathologists; the choice of which biopsy specimens require adjunctive staining showed greater variation.
- Adjunctive p16 IHC showed an overall limited increase in diagnostic agreement, although this may be in part due to the high level of expertise of the 2 pathologists in this study.
- p16 IHC improved diagnostic agreement in cervical biopsy specimens negative for human papillomavirus type 16 but positive for other high-risk types, which may pose greater significance in vaccinated populations.

KEY WORDS

high-grade squamous intraepithelial lesion; p16; interrater reproducibility

Am J Clin Pathol August 2024;161:202-209
HTTPS://DOI.ORG/10.1093/AJCP/AQAE029

Received: December 22, 2023
Accepted: February 23, 2024

Corresponding author: Nicolas Wentzensen; wentzenn@mail.nih.gov

cervical precancer, particularly in HPV-vaccinated populations where types other than HPV 16 are relatively more important.

INTRODUCTION

Despite recent advancements in screening for cervical cancer and vaccination against its causative agent, human papillomavirus (HPV), cervical cancer remains a significant problem for women across the globe. In 2018, cervical cancer was the most common malignancy in 28 countries and leading cause of cancer death among women in 42 countries, most of which were in sub-Saharan Africa and Southeast Asia.[1] The success of cervical cancer screening is based on the ability to detect and treat cervical precancers before invasive cancers occur. Even with cervical cancer prevention efforts in the United States, the rate of advanced-stage cervical cancer diagnoses has shown a modest yearly increase over the past 2 decades.[2] In the United States and many other high-resource settings, diagnosis and management of cervical cancer precursors are based on the assessment of cervical biopsy specimens taken during colposcopy following an abnormal screening test. One interobserver variability study involving 37,486 biopsy specimens showed a wide distribution of benign (5.5%-57.1%), cervical intraepithelial neoplasia (CIN) 1 (23.3%-86.7%), and CIN2 (6.2%-14.4%) diagnoses across 15 laboratories.[3] In addition, the histologic interpretation of cervical biopsy specimens demonstrates significant interobserver variability; in particular, CIN2 has shown poor reproducibility and limited agreement in CIN grading.[4,5] CIN2 is a highly heterogeneous group that includes true precancers and transient HPV infections. Despite this heterogeneity, CIN2 is widely used as the clinical management threshold for treatment in many settings; however, CIN3 is a more reproducible diagnosis among pathologists. CIN3+ is the primary outcome in many natural history and clinical studies, with CIN3+ risks underlying current screening and management guidelines.

Given the importance of histologic assessment for management of cervical precancers, approaches that can improve diagnostic agreement and accuracy among pathologists are important to facilitate detection of true premalignant lesions. Immunohistochemical staining for the p16 protein, a cyclin-dependent kinase inhibitor overexpressed in cells transformed by oncogenic HPV, has displayed high sensitivity and specificity for CIN2 and CIN3.[6] A meta-analysis found that evaluation of p16 positivity in cervical biopsy specimens can improve agreement among community pathologists and increase diagnostic accuracy closer to that of an expert pathologist panel.[7] The Lower Anogenital Squamous Terminology (LAST) project, sponsored by the College of American Pathologists and American Society for Colposcopy and Cervical Pathology, has put forth recommendations for p16 immunohistochemistry (IHC) use as an adjunct to H&E staining in certain situations.[8] However, despite increasing use of the LAST recommendations, the impact on interobserver agreement and accuracy has not been widely studied. In this study, we assessed the impact of p16 IHC on interobserver reproducibility of cervical biopsy specimens between 2 experienced pathologists using 2 forms of evaluation: one based on H&E interpretation alone and graded using the CIN system as originally proposed by Richart[9] and another that included the use of p16 staining to support the H&E diagnosis.

METHODS

Study Population

Cervical biopsy specimens were selected from the Study to Understand Cervical Cancer Early Endpoints and Determinants (SUCCEED); study design and methods have been previously described.[10] Briefly, SUCCEED was a cross-sectional study that sought to elucidate the relationship between HPV genotypes, cofactors, and molecular alterations in cervical carcinogenesis.[11-16] The study was approved by the institutional review boards at the University of Oklahoma Health Sciences Center (OUHSC) and the National Cancer Institute. Between November 2003 and September 2007, women who were referred to OUHSC for colposcopy after an abnormal Papanicolaou test or histologic diagnosis of CIN were approached for study enrollment. Exclusion criteria included age younger than 18 years, pregnancy at time of visit, prior treatment with chemotherapy or radiation for cancer, HIV positivity, previous hysterectomy, and reason for visit to clinic as vaginal-only colposcopy. From the biopsy specimens of over 2000 women who participated in SUCCEED, we randomly targeted formalin-fixed, paraffin-embedded cervical biopsy specimens taken from 300 participants with the following histology and HPV results: 25 normal/benign and HPV negative, 50 normal/benign and HPV positive, 50 CIN1 and HPV positive, 50 CIN2 and HPV positive, 75 CIN3 and HPV positive, and 50 HPV-positive invasive squamous cell carcinomas. Due to inability to retrieve, section, and/or process some biopsy specimens, H&E and p16 IHC slides were generated from 249 of the 300 archived biopsy blocks.

HPV Genotyping

Human papillomavirus genotyping was performed from the residual samples in vials containing PreservCyt solution (Hologic) that were collected for ThinPrep cytology (Hologic) at the time of colposcopy with biopsy. This was done using the LINEAR ARRAY (LA) HPV Genotyping Test (Roche Diagnostics). After vortexing, a 1-mL aliquot from each PreservCyt vial underwent DNA isolation with the QIAmp DNA Blood Mini Kit (Qiagen), as described previously.[17] Isolated DNA was stored at –70°C until amplification by polymerase chain reaction was performed with the LA HPV Genotyping Test. The LA assay detects up to 37 HPV genotypes (6, 11, 16, 18, 26, 31, 33, 35, 39, 40, 42, 45, 51, 52, 53, 54, 55, 56, 58, 59, 61, 62, 64, 66, 67, 68, 69, 70, 71, 72, 73, 81, 82, 83, 84, IS39, CP6108). As an internal control ensuring adequate amplifiable DNA in each specimen, 2 different concentrations of β-globin probes were present on each strip. Positive and negative controls provided by the manufacturer were processed with each run of DNA isolation, amplification, and detection of specimens. Including controls, up to 84 specimens were amplified in 1 run using a 10-µL template DNA. Up to 30 specimens underwent hybridization to linear arrays at one time, and signals were detected with the Auto-LiPA instrument

HTTPS://DOI.ORG/10.1093/AJCP/AQAE029

(Innogenetics). We considered the following HPV genotypes high risk for causing cervical cancer in women: HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, and 59.[18]

Slide Preparation

Two adjacent sections were cut from each archived biopsy block and mounted onto glass slides: one for H&E staining and the other for p16 IHC using the CINtec Histology Kit (Roche mtm laboratories), according to instructions provided by the manufacturer. Briefly, antigen retrieval was performed by soaking the slides in a 95°C to 99°C water bath for 10 minutes. Endogenous peroxidase activity was blocked, and slides were then incubated with the p16 antibody (clone E6H4) or negative reagent control (isotype control antibody) for 30 minutes. A secondary antibody reagent (polymer-based goat–anti-mouse antibody fragment conjugated to horseradish peroxidase) was applied for 30 minutes. After visualization with 3,3'-diaminobenzidine chromogen and counterstaining with hematoxylin, coverslips were placed on each glass slide. For each run, a positive control slide containing tissue from a cervical biopsy specimen with known positive immunoreactivity for p16 was used to ensure proper staining technique.

Slide Interpretation

At the time of the SUCCEED study, histologic interpretation of biopsy and loop electrosurgical excision procedure specimens was performed by the study pathologist (R.Z.) at OUHSC using CIN terminology; these diagnoses were subsequently used for clinical management. For this study, newly cut and processed slides (1 H&E stained and 1 immunostained for p16) were evaluated by 2 pathologists (R.Z. and T.M.D.); each had access to information regarding each patient's age and referral cytology result. Each pathologist first evaluated the H&E-stained slides alone and assigned diagnoses based on the following options: normal or benign, atypical squamous metaplasia, CIN1, CIN2, CIN3, squamous cancer, adenocarcinoma in situ, adenocarcinoma, or other. Each pathologist then indicated if they would order p16 IHC to facilitate diagnosis. After a washout period of at least 2 weeks, each pathologist performed an additional evaluation of the paired H&E-stained slides and p16 slides, regardless of whether or not they requested p16 IHC at the first time point. The LAST terminology was used for squamous intraepithelial lesions in the setting of H&E plus p16. A p16 stain was considered positive if there was continuous staining of cells (nuclei and cytoplasm) in the basal and parabasal layers for at least one-third of the thickness of the squamous epithelium, regardless of the staining pattern in the more superficial layers. A slide was considered negative for p16 immunoreactivity if staining was noncontinuous (ie, focal or patchy), if staining was not present in the basal and parabasal layers, or if the biopsy specimen did not display immunoreactivity anywhere along the squamous epithelium.

Statistical Analysis

Percentage agreement and κ values were calculated to assess interobserver reproducibility of both H&E only and H&E plus p16 interpretation. Confidence intervals were reported according to the bootstrap method.[19] Biopsy specimens given an "other" diagnosis were reviewed in conjunction with their accompanying comments, discussed with the pathologist, and recategorized according to the categories listed above. Five biopsy specimens that were determined to be insufficient for proper visualization of the epithelium by one or both pathologists were dropped from the current study, bringing the total number of biopsy specimens included in this analysis to 244. For assessment of interrater agreement, biopsy specimens with diagnoses of squamous cell carcinomas, adenocarcinomas, and other cancers (including adenosquamous carcinoma and small cell neuroendocrine cancer of the cervix) were combined into one general "cancer" category.

For analyses primarily focused on distinguishing precancerous from nonprecancerous lesions (ie, less than high-grade squamous intraepithelial lesion [HSIL] vs HSIL+), cancer cases were excluded, leaving a total of 200 biopsy specimens for comparison. In additional analyses, we calculated agreement statistics after removing p16 evaluation from biopsy specimens for which the reviewing pathologist did not request p16 IHC and, for comparison purposes, reclassified their H&E-only diagnoses with LAST nomenclature: CIN1 without p16 IHC was renamed low-grade squamous intraepithelial lesion (LSIL), CIN2 without p16 IHC was renamed HSIL-CIN2, and CIN3 without p16 IHC was renamed HSIL-CIN3. To account for cases where the 2 pathologists differed in their decision to order p16 IHC, we also considered agreement values at the individual biopsy level; if either pathologist determined p16 IHC was necessary, the H&E plus p16 IHC interpretations made at the second time point for both pathologists were compared. If neither pathologist ordered p16 IHC, their H&E-only diagnoses were reclassified using LAST nomenclature as detailed above and compared.

RESULTS

Study Population

The 244 study participants had a median age of 26 years and a mean age of 31 years (range, 18-81 years). Using clinical histology diagnoses previously made during the SUCCEED study, there were 55 (22.5%) negative, 37 (15.2%) CIN1, 44 (18.0%) CIN2, 64 (26.2%) CIN3, and 44 (18.0%) invasive carcinoma biopsy specimens. Most women were positive for HPV 16 (55%). TABLE 1 shows a summary of participant age and HPV genotype positivity, stratified by prior clinical histology result in the SUCCEED study. Per study design, most cases selected were from participants who were HPV positive, including for multiple HPV genotypes (57.4%). Eighteen of 55 histologically negative/benign biopsy specimens were negative for HPV, and there were no cases that were only positive for low-risk HPV genotypes. We assessed HPV positivity based on positivity for HPV 16, then positivity for HPV 18 and/or HPV 45; these genotypes are most common in both high-grade intraepithelial lesions and invasive cervical cancers and are included in current routine HPV vaccines available in the United States. As such, pathologists may be most familiar with the morphologic alterations associated with these HPV genotypes. Thus, interrater agreement of cases positive for high-risk genotypes

HTTPS://DOI.ORG/10.1093/AJCP/AQAE029

TABLE 1 Study Population Characteristics by Clinical Diagnosis[a]

Determinants	Prior Clinical Histology Results					
	Negative (n = 55)	CIN1 (n = 37)	CIN2 (n = 44)	CIN3 (n = 64)	Invasive squamous carcinoma (n = 44)	Total (n = 244)
Mean age at biopsy, y	30	25	26	29	46	31
HPV genotype positivity, No. (%)						
HPV 16+	19 (35)	11 (30)	21 (48)	51 (80)	32 (73)	134 (55)
HPV 16–; HPV 18 and/or 45+	6 (11)	7 (19)	3 (7)	4 (6)	4 (9)	24 (10)
HPV 16, 18, 45–; other high-risk HPV+[b]	12 (22)	19 (51)	20 (45)	9 (14)	8 (18)	68 (27)
Negative for all HPV genotypes	18 (33)	0 (0)	0 (0)	0 (0)	0 (0)	18 (7)

CIN, cervical intraepithelial neoplasia; HPV, human papillomavirus.
[a]In some instances, percentages do not add to 100% due to missing information from the database regarding preceding cytology for 5 biopsy specimens and HPV genotype for 1 biopsy specimen.
[b]Other high-risk HPV genotypes include HPV 31, 33, 35, 39, 51, 52, 56, 58, 59, and/or 68.

TABLE 2 Summary of Indicated Reasons for Ordering p16 Immunohistochemistry by Pathologist[a]

Characteristic	Total biopsies, No. (%)	
	Pathologist 1	Pathologist 2
Precancer vs mimic (LAST recommendation 1)	47 (23.5)	27 (13.5)
H&E CIN2 (LAST recommendation 2)	15 (7.5)	9 (4.5)
Professional disagreement (LAST recommendation 3)	0 (0)	0 (0)
High-risk colposcopic referral (LAST recommendation 4a)	0 (0)	2 (1)
Other[b]	0 (0)	2 (1)
Total	62 (31)	40 (20)

CIN, cervical intraepithelial neoplasia; LAST, Lower Anogenital Squamous Terminology.
[a]For all biopsies that the reviewing pathologist requested to order p16 immunohistochemistry for further evaluation, the indicated reasons, corresponding to recommendations for p16 use, as outlined by the LAST project,[8] are listed. Estimated percentages for each reason were proposed by the LAST project and are as following: precancer vs mimic (10%), H&E CIN2 (1%), professional disagreement (1%), and high-risk colposcopic referral (3%). This results in suggested adjunctive p16 immunohistochemistry (IHC) utilization for less than 25% of all cervical biopsy specimens. Professional disagreement as a reason to order p16 IHC could not be accounted for in this study due to its design.
[b]The indicated reason was not one of the LAST recommendations.

other than HPV 16 was of particular interest, since high-grade lesions caused by these genotypes may predominate in HPV-vaccinated cohorts,[20] resulting in increased interobserver variation in biopsy diagnoses.

Comparison of p16 Assessment and Indications for Ordering

Between the 2 pathologists, there was excellent agreement (92.5%, κ = 0.85; 95% CI, 0.78-0.92) in the evaluation of p16 IHC as positive or negative. The choice of which biopsies needed p16 IHC, however, displayed greater variability, with a percentage agreement of 78.0% and κ value of 0.43 (95% CI, 0.29-0.57). TABLE 2 shows the percentage of biopsy specimens with p16 IHC requested on initial H&E review and the indicated reason for each pathologist. Both pathologists requested p16 IHC for all H&E biopsy specimens with a diagnosis of atypical squamous metaplasia. For almost all these cases, the indicated reason was "precancer vs mimic," consistent with LAST recommendation 1. Pathologist 2 requested p16 IHC on 2 biopsy specimens with HSIL on preceding cytology, consistent with LAST recommendation 4a; both were also interpreted as atypical squamous metaplasia on initial H&E review. While pathologist 2 requested p16 for all 9 biopsy specimens with a diagnosis of CIN2 on H&E alone, consistent with LAST recommendation 2, pathologist 1 requested p16 for a subset of CIN2 H&E biopsy specimens (10 of 35). Twenty-two of the 25 CIN2 biopsy specimens for which p16 was not requested by pathologist 1 were diagnosed as CIN3 by pathologist 2. This finding is likely due to differences in the threshold of CIN2 vs CIN3 between the 2 pathologists and reflects the larger heterogeneity in assessing CIN2 among pathologists.

Interobserver Agreement of H&E Only vs H&E Plus p16 Diagnoses

Results of H&E only and H&E plus p16 diagnoses are shown in TABLES 3 and 4, respectively. Agreement statistics are detailed in TABLE 5 and summarized here. The 2 pathologists' diagnoses were compared before and after they were dichotomized into 2 categories based on the generally accepted threshold in treatment decision-making: (1) atypical squamous metaplasia and benign vs CIN2 and worse for H&E-only evaluation and (2) LSIL and benign vs HSIL and worse for H&E plus p16 evaluation. Percentage agreement values in diagnoses ranged from 65.5% to 70.0%; this increased to 90.0% to 92.5% when diagnoses were dichotomized. Nonweighted κ values ranged from 0.54 to 0.58 for the expanded categorization and 0.79 to 0.85 for the

HTTPS://DOI.ORG/10.1093/AJCP/AQAE029

TABLE 3 Reproducibility of H&E-Only Evaluation[a]

	Pathologist 2						
Pathologist 1	Normal/benign, No.	CIN1, No.	ASM, No.	CIN2, No.	CIN3, No.	Cancer, No.	Total No.
Normal/benign	**61**	2	4	0	0	0	67
CIN1	4	**19**	2	2	2	0	29
ASM	8	2	**20**	4	4	0	38
CIN2	1	3	1	**3**	27	0	35
CIN3	1	0	1	0	**25**	1	28
Cancer	0	0	0	0	0	**3**	3
Total	75	26	28	9	58	4	200

ASM, atypical squamous metaplasia; CIN, cervical intraepithelial neoplasia.
[a]Comparisons of H&E-only diagnoses are shown. Values that represent diagnostic agreement between the 2 pathologists are in bold. Because CIN2 is employed as the threshold in the treatment vs surveillance decision in most cases, any values that represent agreement in the less than CIN2 vs CIN2+ dichotomy are shaded in green, whereas those that represent disagreement are shaded in orange.

TABLE 4 Reproducibility of H&E Plus p16 Evaluation[a]

	Pathologist 2					
Pathologist 1	Normal/benign, No.	LSIL, No.	HSIL + HSIL-CIN2, No.	HSIL-CIN3, No.	Cancer, No.	Total No.
Normal/benign	**79**	2	2	1	0	84
LSIL	5	**13**	0	0	0	18
HSIL + HSIL-CIN2	1	10	**14**	34	0	59
HSIL-CIN3 + AIS	2	0	1	**32**	1	36
Cancer	0	0	0	1	**2**	3
Total	87	25	17	68	3	200

AIS, adenocarcinoma in situ; CIN, cervical intraepithelial neoplasia; HSIL, high-grade squamous intraepithelial lesion; LSIL, low-grade squamous intraepithelial lesion.
[a]Comparisons of H&E plus p16 evaluation diagnoses are shown. Values that represent diagnostic agreement between the 2 pathologists are in bold. Because CIN2 is employed as the threshold in the treatment vs surveillance decision in most cases, any values that represent agreement in the less than CIN2 vs CIN2+ dichotomy are shaded in green, whereas those that represent disagreement are shaded in orange.

TABLE 5 Percentage Agreement and κ Values by Scoring System

Scoring system	Diagnosis categorization	Percentage agreement	Nonweighted κ (95% CI)	Linear-weighted κ (95% CI)
H&E only	Six-level[a]	65.5	0.56 (0.51-0.59)	0.72 (0.70-0.78)
	Dichotomized[b]	90.5	0.79 (0.70-0.88)	
H&E plus p16 IHC when ordered by reviewing pathologist	Five-level[c]	66.5	0.54 (0.49-0.62)	0.72 (0.68-0.73)
	Dichotomized[d]	90.0	0.80 (0.71-0.88)	
H&E plus p16 IHC when ordered by at least 1 pathologist	Five-level	67.5	0.55 (0.51-0.60)	0.73 (0.68-0.75)
	Dichotomized	92.5	0.85 (0.78-0.92)	
H&E plus p16 IHC for all biopsy specimens	Five-level	70.0	0.58 (0.49-0.62)	0.76 (0.73-0.78)
	Dichotomized	92.0	0.84 (0.77-0.92)	

CIN, cervical intraepithelial neoplasia; HSIL, high-grade squamous intraepithelial lesion; LAST, Lower Anogenital Squamous Terminology; LSIL, low-grade squamous intraepithelial lesion.
[a]H&E-only 6-level categorization refers to the following grouping: normal/benign vs atypical squamous metaplasia vs CIN1 vs CIN2 vs CIN3 vs cancer.
[b]H&E-only dichotomized categorization refers to the following grouping: normal/benign and CIN1 and atypical squamous metaplasia vs CIN2 and CIN3 and cancer.
[c]H&E plus p16 five-level categorization refers to the following grouping: normal/benign vs LSIL vs HSIL and HSIL-CIN2 vs HSIL-CIN3 vs cancer. H&E biopsy specimens without adjunctive p16 evaluation were renamed with LAST-concordant nomenclature, as described in the Materials and Methods section.
[d]H&E plus p16 dichotomized categorization refers to the following grouping: normal/benign and LSIL vs HSIL and HSIL-CIN2 and HSIL-CIN3 and cancer.

dichotomized categorization. Linear-weighted κ values ranged from 0.72 to 0.76.

Adjunctive p16 IHC evaluation when ordered by the reviewing pathologist showed only a minimal change in interrater agreement from H&E-only evaluation (nonweighted κ values of 0.56-0.54); this effect was preserved when considering the more clinically relevant dichotomized categorization (0.79-0.80). In the scenario in which adjunctive p16 evaluation was used for the biopsy specimen when at least 1 pathologist ordered it, however, the dichotomous agreement increased (0.79-0.85). Similarly, when p16 IHC was used for all biopsy specimens, regardless if the pathologist would request it or not, dichotomous agreement increased (0.79-0.84), even though improvement in agreement of actual diagnoses was less substantial (0.56-0.58). Six biopsy

HTTPS://DOI.ORG/10.1093/AJCP/AQAE029

TABLE 6 The κ Values Stratified by HPV Genotype and Scoring System[a]

Characteristic	Scoring system, κ (95% CI)			
	H&E only	H&E plus p16 IHC when ordered by reviewing pathologist	H&E plus p16 IHC when ordered by at least 1 pathologist	H&E plus p16 IHC for all biopsy specimens
Total (n = 200)	0.79 (0.70-0.88)	0.80 (0.71-0.88)	0.85 (0.78-0.92)	0.84 (0.77-0.92)
HPV 16+ (n = 102)	0.80 (0.69-0.92)	0.78 (0.65-0.90)	0.80 (0.68-0.92)	0.80 (0.68-0.92)
HPV 16–; HPV 18 and/or 45+ (n = 20)	0.86 (0.59-1.00)	0.89 (0.69-1.00)	1.00	0.90 (0.70-1.00)
HPV 16, 18, 45–; other HRHPV+ (n = 60)	0.58 (0.32-0.84)	0.71 (0.52-0.90)	0.82 (0.68-0.97)	0.82 (0.68-0.97)

HPV, human papillomavirus; HRHPV, high-risk human papillomavirus.
[a]The κ values are based on dichotomized diagnosis categorization (lower than low-grade squamous intraepithelial lesion vs higher than high-grade squamous intraepithelial lesion).

TABLE 7 Reproducibility of H&E Plus p16 Evaluation for Biopsy Specimens Positive for High-Risk HPV Genotypes Other Than HPV 16, 18, and/or 45[a]

Pathologist 1	Pathologist 2				
	Normal/benign, No.	LSIL, No.	HSIL-CIN2, No.	HSIL-CIN3, No.	Total No.
Normal/benign	**24**	1	1	0	26
LSIL	3	**7**	0	0	10
HSIL + HSIL-CIN2	0	3	**7**	7	17
HSIL-CIN3	1	0	0	**6**	7
Total	28	11	8	13	60

CIN, cervical intraepithelial neoplasia; HSIL, high-grade squamous intraepithelial lesion; LSIL, low-grade squamous intraepithelial lesion.
[a]High-risk human papillomavirus (HPV) genotypes represented in these biopsy specimens include HPV 31, 33, 35, 39, 51, 52, 56, 58, and 59. Values that represent diagnostic agreement between the 2 pathologists are in bold. Because CIN2 is employed as the threshold in the treatment vs surveillance decision in most cases, any values that represent agreement in the less than CIN2 vs CIN2+ dichotomy are shaded in green, whereas those that represent disagreement are shaded in orange.

specimens received diagnoses that were extremely discordant between the 2 pathologists (ie, determined by one pathologist as normal/benign and HSIL or higher grade by the other). Five of these cases were due to disagreements in evaluating the adjunctive p16 slide as positive vs negative.

Interobserver Agreement Stratified by HPV Genotype

The κ values were calculated among groups stratified by HPV genotype category. Results are summarized in TABLE 6. Biopsy specimens taken from women who tested positive for HPV 16 showed similar agreement between H&E only and H&E plus p16 evaluation (κ = 0.80 vs 0.78-0.80). Biopsy specimens taken from women who tested negative for HPV 16 but positive for HPV 18 and/or 45 showed overall higher agreement than the nonstratified set, with κ values ranging from 0.86 to 1.00, compared with 0.79 to 0.85. Notably, in these cases, interrater agreement reached 100% when p16 IHC was used when requested by one or both pathologists. However, given the small sample size of 20 biopsy specimens, this finding should be approached with caution. The largest improvement in agreement was demonstrated when p16 evaluation was added for biopsy specimens collected from women who were negative for HPV 16, 18, and 45 but positive for other high-risk strains, with improvements in κ from 0.58 to 0.71 to 0.82. TABLE 7 compares the H&E plus p16 diagnoses between the 2 pathologists for this set of biopsy specimens. Thirty-two biopsy specimens received differing diagnoses between the 2 pathologists based on H&E-only evaluation that then received the same diagnosis with adjunctive p16 IHC. Seventeen of these biopsy specimens were positive for HPV 16, 18, and/or 45, and 2 were HPV negative. Among the 11 remaining biopsy specimens, the following high-risk HPV genotypes were represented, in decreasing order: HPV 33, 35, 52, and 58 (present in 5 biopsy specimens); HPV 39 and 59 (4 biopsy specimens); HPV 31 and 51 (3 biopsy specimens); and HPV 56 (1 biopsy specimen).

DISCUSSION

We investigated the impact of p16 IHC on reproducibility of cervical histology evaluation among 2 expert pathologists. Evaluation of p16 itself (positive vs negative) was highly reproducible, with a κ value of 0.85. The decision of whether to order p16 was more heterogeneous between the 2 pathologists, with a κ value of 0.43, likely demonstrating differences in approach to p16 IHC itself and diagnosing CIN2. Overall, we found a limited increase in agreement when the 2 pathologists used p16 IHC for all biopsy specimens in the set. However, there was a marked increase in agreement when adjunctive p16 evaluation was used for biopsy specimens positive for high-risk genotypes other than HPV 16.

There was significant variation between the 2 pathologists in the use of CIN2 as a diagnostic category. On H&E alone, out of 56 "high-grade" (diagnosed by either pathologist as CIN2 or CIN3) biopsy specimens, pathologist 1 diagnosed 35 CIN2, whereas pathologist 2 diagnosed 9 CIN2. Nearly half (27) of the 58 biopsy specimens that pathologist 2 classified as CIN3 were classified as CIN2 by pathologist 1. Twenty-five of these 27 cases were p16 positive per pathologist 1, thus supporting a diagnosis of HSIL. No biopsy specimens were classified as CIN3 by pathologist 1 and

HTTPS://DOI.ORG/10.1093/AJCP/AQAE029

CIN2 by pathologist 2. Of the 46 biopsy specimens that were given a diagnosis of atypical squamous metaplasia based on H&E-only evaluation by a pathologist, there was diagnostic agreement (ie, both pathologists assigned a diagnosis of atypical squamous metaplasia) for 20 (43.5%) of them. Twenty-one and 19 of these biopsy specimens were determined by both pathologists to be negative and positive for p16 IHC, respectively. With adjunctive p16 IHC evaluation, diagnostic agreement increased to 24 (52.2%) biopsy specimens, with 7 biopsy specimens that were mutually upgraded to HSIL(-CIN2) and 17 that were mutually downgraded to normal/benign. While the LAST project recommends specific situations to use p16 to clarify uncertain diagnoses and to support a diagnosis of HSIL(-CIN2), our data show that among 2 expert pathologists, the distinction between grading a lesion as CIN2 vs CIN3 was not relevant, as most CIN2 biopsy specimens were determined by both to be p16 positive. Because both of these diagnoses fall under the HSIL category, this distinction had little impact on the dichotomous comparisons and may have minimal impact on clinical management when CIN2 is used as the threshold for treatment. However, some centers prefer the option of observing CIN2 in young women who desire future pregnancies, particularly when HPV testing is included, highlighting the continued need to separate CIN3 and CIN2 from HSIL. Furthermore, CIN2 diagnoses downgraded by a negative adjunctive p16 result to LSIL highlights the benefit of p16 IHC to reduce unnecessary harm; in our study, this occurred for 4 of 35 CIN2 biopsy specimens by pathologist 1 and 0 of 9 CIN2 biopsy specimens by pathologist 2. Interrater agreement did not increase when p16 evaluation was used as requested by individual pathologists compared with H&E-only evaluation but did improve when p16 evaluation was used for all biopsy specimens, similar to the results of the 2010 study by Bergeron et al,[21] which demonstrated improved diagnostic consistency of community pathologists to each other and to an expert gynecologic pathologist panel. The differences in interrater agreement in this study are not large; because the pathologists included in this study are experts in gynecological pathology, we hypothesize that improvements produced by adjunctive p16 evaluation may not have been as sizable as in a community setting.

When stratifying results based on HPV genotype, biopsy specimens taken from women who were HPV 16 positive did not show a sizable increase in reproducibility with adjunctive p16 evaluation, suggesting these precancers are easy to diagnose based on morphology alone. In contrast, interrater agreement for HPV 16–negative and HPV 18–positive and/or HPV 45–positive biopsy specimens increased for all scoring systems, with the greatest gain when it was ordered by one or both pathologists. This finding, however, is limited by a small sample size of 20 biopsy specimens. The largest increase in agreement when p16 IHC supplemented H&E diagnostic evaluation occurred in biopsy specimens negative for HPV 16, 18, and 45 but positive for other high-risk HPV types, with an increase in κ from 0.58 to 0.82.

With the introduction of HPV vaccination, prevalence of HSIL associated with HPV 16 has decreased.[22] Given that HPV 16 is the most frequent causative oncogenic HPV genotype, if it continues to decrease in prevalence, accurate identification of precancers caused by non–HPV 16 types will become relatively more clinically relevant.

Further studies comparing the role of p16 immunostaining among various HPV genotypes are needed to make further conclusions about how to better detect HSIL caused by nonvaccine genotypes. Our next steps are to continue these efforts to investigate p16 performance among specific HPV genotypes, as well as establish gold standard diagnoses of biopsy specimens to examine the role of p16 immunostaining in increasing not only the reproducibility but also the accuracy of cervical histopathology. The high reproducibility of p16 is encouraging, as this supports the use of a biomarker as an objective measure in improving the practice of cervical biopsy interpretation; however, as highlighted by the several biopsy specimens that had extremely discordant diagnoses (classified as normal/benign by one pathologist and HSIL or higher grade by the other), p16 interpretation remains somewhat subjective. It remains important to elucidate the situations where p16 immunostaining, possibly in conjunction with additional biomarkers, will accurately signal which high-grade lesions are appropriate to treat while avoiding the overtreatment of lesions that would be unlikely to progress to cancer with simple surveillance.

Acknowledgments: This work was supported by the Intramural Research Program at the National Cancer Institute of the National Institutes of Health (Z01 CP010124).

Conflict of interest disclosure: The authors have nothing to disclose.

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68(6):394-424. https://doi.org/10.3322/caac.21492
2. Francoeur AA, Liao CI, Casear MA, et al. The increasing incidence of stage IV cervical cancer in the USA: what factors are related? *Int J Gynecol Cancer.* 2022;32:1115-1122.
3. Gage JC, Schiffman M, Hunt WC, et al; New Mexico HPV Pap Registry Steering Committee. Cervical histopathology variability among laboratories: a population-based statewide investigation. *Am J Clin Pathol.* 2013;139(3):330-335. https://doi.org/10.1309/AJCPSD3ZXJXP7NNB
4. Carreon JD, Sherman ME, Guillen D, et al. CIN2 is a much less reproducible and less valid diagnosis than CIN3: results from a histological review of population-based cervical samples. *Int J Gynecol Pathol.* 2007;26(4):441-446. https://doi.org/10.1097/pgp.0b013e31805152ab
5. Stoler MH, Schiffman M; Atypical Squamous Cells of Undetermined Significance-Low-Grade Squamous Intraepithelial Lesion Triage Study (ALTS) Group. Interobserver reproducibility of cervical cytologic and histologic interpretations: realistic estimates from the ASCUS-LSIL Triage Study. *JAMA.* 2001;285(11):1500-1505. https://doi.org/10.1001/jama.285.11.1500
6. Galgano MT, Castle PE, Atkins KA, Brix WK, Nassau SR, Stoler MH. Using biomarkers as objective standards in the diagnosis of cervical biopsies. *Am J Surg Pathol.* 2010;34(8):1077-1087. https://doi.org/10.1097/PAS.0b013e3181e8b2c4
7. Reuschenbach M, Wentzensen N, Dijkstra MG, von Knebel Doeberitz M, Arbyn M. p16INK4a immunohistochemistry in cervical biopsy specimens: a systematic review and meta-analysis of the interobserver agreement. *Am J Clin Pathol.* 2014;142(6):767-772. https://doi.org/10.1309/AJCP3TPHV4TRIZEK

8. Darragh TM, Colgan TJ, Cox JT, et al; Members of LAST Project Work Groups. The lower anogenital squamous terminology standardization project for HPV-associated lesions: background and consensus recommendations from the College of American Pathologists and the American Society for Colposcopy and Cervical Pathology. *J Low Genit Tract Dis.* 2012;16(3):205-242. https://doi.org/10.1097/LGT.0b013e31825c31dd
9. Richart RM. Natural history of cervical intraepithelial neoplasia. *Clin Obstet Gynecol.* 1967;10(4):748-784. https://doi.org/10.1097/00003081-196712000-00002
10. Wang SS, Zuna RE, Wentzensen N, et al. Human papillomavirus cofactors by disease progression and human papillomavirus types in the study to understand cervical cancer early endpoints and determinants. *Cancer Epidemiol Biomarkers Prev.* 2009;18(1):113-120. https://doi.org/10.1158/1055-9965.EPI-08-0591
11. Clarke MA, Luhn P, Gage JC, et al. Discovery and validation of candidate host DNA methylation markers for detection of cervical precancer and cancer. *Int J Cancer.* 2017;141(4):701-710. https://doi.org/10.1002/ijc.30781
12. Bodelon C, Vinokurova S, Sampson JN, et al. Chromosomal copy number alterations and HPV integration in cervical precancer and invasive cancer. *Carcinogenesis.* 2016;37(2):188-196. https://doi.org/10.1093/carcin/bgv171
13. Wentzensen N, Walker J, Schiffman M, et al. Heterogeneity of high-grade cervical intraepithelial neoplasia related to HPV16: implications for natural history and management. *Int J Cancer.* 2013;132(1):148-154. https://doi.org/10.1002/ijc.27577
14. Wentzensen N, Schiffman M, Dunn ST, et al. Grading the severity of cervical neoplasia based on combined histopathology, cytopathology, and HPV genotype distribution among 1,700 women referred to colposcopy in Oklahoma. *Int J Cancer.* 2009;124(4):964-969. https://doi.org/10.1002/ijc.23969
15. Wentzensen N, Schiffman M, Dunn T, et al. Multiple human papillomavirus genotype infections in cervical cancer progression in the study to understand cervical cancer early endpoints and determinants. *Int J Cancer.* 2009;125(9):2151-2158. https://doi.org/10.1002/ijc.24528
16. Yang HP, Zuna RE, Schiffman M, et al. Clinical and pathological heterogeneity of cervical intraepithelial neoplasia grade 3. *PLoS One.* 2012;7(1):e29051. https://doi.org/10.1371/journal.pone.0029051
17. Dunn ST, Allen RA, Wang S, Walker J, Schiffman M. DNA extraction: an understudied and important aspect of HPV genotyping using PCR-based methods. *J Virol Methods.* 2007;143(1):45-54. https://doi.org/10.1016/j.jviromet.2007.02.006
18. Bouvard V, Baan R, Straif K, et al; WHO International Agency for Research on Cancer Monograph Working Group. A review of human carcinogens—part B: biological agents. *Lancet Oncol.* 2009;10(4):321-322. https://doi.org/10.1016/s1470-2045(09)70096-8
19. Efron B. *The Jackknife, the Bootstrap, and Other Resampling Plans.* Society for Industrial and Applied Mathematics; 1982:92.
20. de Sanjose S, Quint WG, Alemany L, et al; Retrospective International Survey and HPV Time Trends Study Group. Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study. *Lancet Oncol.* 2010;11(11):1048-1056. https://doi.org/10.1016/S1470-2045(10)70230-8
21. Bergeron C, Ordi J, Schmidt D, Trunk MJ, Keller T, Ridder R; European CINtec Histology Study Group. Conjunctive p16INK4a testing significantly increases accuracy in diagnosing high-grade cervical intraepithelial neoplasia. *Am J Clin Pathol.* 2010;133(3):395-406. https://doi.org/10.1309/AJCPXSVCDZ3D5MZM
22. Giannella L, Delli Carpini G, Di Giuseppe J, et al. Trend of HPV 16/18 genotypes in cervical intraepithelial neoplasia grade 3: data for 2007-2018. *Infect Drug Resist.* 2021;14:3763-3771. https://doi.org/10.2147/IDR.S326851

HOLOGIC®

First and Only FDA Cleared
Digital Cytology System

Genius™ Cervical AI

Genius™ Review Station

Genius™ Digital Imager

Empower Your Genius With Ours

Make a Greater Impact on Cervical Cancer
with the Advanced Technology of the
Genius™ Digital Diagnostics System

Click or Scan
to discover more

genius™
DIGITAL DIAGNOSTICS