Cervical Histopathology Variability Among Laboratories

A Population-Based Statewide Investigation

Julia C. Gage, PhD, MPH,1 Mark Schiffman, MD, MPH,1 William C. Hunt, MA,2 Nancy Joste, MD,2 Arpita Ghosh, PhD,1 Nicolas Wentzensen, MD, PhD,1 and Cosette M. Wheeler, PhD,2 for The New Mexico HPV Pap Registry Steering Committee

Key Words: Cervical intraepithelial neoplasia; Histopathology; Reproducibility

DOI: 10.1309/AJCPSD3ZXJXP7NNB

Abstract

To inform the proposed systematic adjudicative staining of cervical intraepithelial neoplasia grade 2 (CIN2) and equivocal diagnoses, we characterized diagnostic heterogeneity across 15 laboratories. Laboratory-specific distributions of 37,486 biopsy specimen diagnoses were compared after adjusting for preceding cytology. In a subset of preceding cytology specimens, HPV16 genotyping was considered an indicator of lesion severity. Distributions of normal and CIN1 diagnoses varied widely, with laboratories favoring either normal (5.5%-57.7%) or CIN1 diagnoses (23.3%-86.7%; P < .001 for normal:CIN1 variability). Excluding extreme values, 6.2% to 14.4% of diagnoses were CIN2 (P < .001). For CIN2 diagnoses, HPV16 positivity in the preceding cytology varied between 39.0% in the largest laboratory and 57.4% in others (P < .001), suggesting differential interpretation, not population differences, as a cause of variability. In conclusion, the frequency of diagnoses requiring special staining (p16INK4a immunostaining) to adjudicate equivocal CIN2 will be sizable and vary between laboratories, especially if extended to a fraction of CIN1 lesions.

Histopathology is the reference standard for diagnosing cervical intraepithelial neoplasia (CIN) and informs clinical management by identifying which women will be treated, followed, or returned to routine screening. In addition, cervical histopathology defines the endpoints in epidemiological studies of cervical carcinogenesis and accuracy of screening tests, as well as population-based surveillance studies on the impact of human papillomavirus (HPV) vaccination.

Comparisons between pathologists have found considerable variability in histology grading of CIN.1-9 Typically, CIN grade 3 (CIN3) has been considered definitive precancer, with an approximately 30% risk of progression to cervical cancer,10 whereas CIN grade 2 (CIN2) is the conservative threshold for treatment in clinical practice. More recently, the diagnosis of CIN2 has been contentious, with the notion that CIN2 might not represent a distinct disease state but rather an equivocal, heterogeneous category consisting of benign HPV infection and precancer.8 Similar to a cytologic classification of high-grade squamous intraepithelial neoplasia (HSIL), some pathologists favor combining CIN2 and CIN3.11-13 Furthermore, the classification of CIN1 vs normal is also variable.3 Some studies suggest that p16INK4a immunostaining in histologic specimens can be used to classify CIN2 and equivocal CIN1-CIN2 diagnoses as either high or low grade.14,15

If histopathology classifications vary to the extent reported in pairwise comparisons of pathologists, such variability might be observed across laboratories. Ideally, if histopathology were an accurate and reproducible measure of cervical neoplasia, the same diagnosis would be rendered for any given case regardless of which pathologist received the specimen. The choice of laboratory would not matter; for any given group of cases, the same distribution of diagnoses
would be rendered. We determined whether 15 histopathology laboratories have similar distributions of normal, CIN1, CIN2, and CIN3 diagnoses after accounting for variability between patient populations. We measured the variability in CIN2 and roughly how many cases would be recommended for adjunctive staining due to CIN2 or equivocal CIN2.

Materials and Methods

We analyzed pathology diagnoses for cervical biopsy specimens read between January 2006 and June 2011 at 15 laboratories using data from the New Mexico HPV Pap Registry (NMHPVPR). The NMHPVPR is a public health surveillance activity established to evaluate the continuum of cervical cancer prevention throughout the state. All Papanicolaou (Pap) and HPV tests and all cervical, vulvar, and vaginal pathology results are reportable under the New Mexico Notifiable Diseases and Conditions (http://nmhealth.org/ERD/healthdata/documents/NotifiableDiseasesConditions022912final.pdf).

For patients with multiple biopsy specimens taken at the same clinic visit, the biopsy specimen with the worst diagnosis was considered in this analysis. Neither loop electrosurgical excision procedure nor endocervical curettage specimens were included. Because not all diagnoses were categorized as normal, CIN1, CIN2, CIN3, or carcinoma, the following combinations were made: diagnoses of koilocytosis and HPV effects were categorized as CIN1, the uncommon diagnosis of CIN1-2 was categorized as CIN2, and the common diagnosis of CIN2-3 was kept as a separate category. The diagnosis of carcinoma in situ was categorized as CIN3. Finally, the uncommon diagnosis of HSIL, without any more specific designation, was categorized as CIN2-3. We compared distributions of CIN2-3 relative to CIN2 and CIN3 to determine whether such combinations might reduce laboratory variability.

Because some laboratories might have a patient population with more HPV-related cervical disease, we considered the patient’s age and the severity of the preceding Pap smear cytology result that led to the biopsy if taken within 1 year of the biopsy specimen being processed. Preceding cytology results were available for 89% of the biopsy specimens. The mean age and severity of the preceding cytology were compared among laboratories with $F$ tests and likelihood ratio $\chi^2$ tests, respectively.

The proportions of diagnoses (normal, CIN1, CIN2, CIN2-3, CIN3, and carcinoma) were compared across laboratories with the likelihood ratio $\chi^2$ test. To adjust for patient variability between laboratories, biopsy specimen results were directly standardized to the cytology and age distributions of the laboratory reading the greatest number of biopsy specimens. For standardization, cytology results were combined into 3 categories: (1) no preceding cytology within the past year, (2) normal/atypical squamous cells of undetermined significance (ASC-US)/low-grade squamous intraepithelial lesions (LSIL), and (3) atypical squamous cells, cannot exclude high-grade (ASC-H)/atypical glandular cells of undetermined significance (AGUS)/HSIL or worse. Age was classified as younger than 25 years, 25 to 39 years, and 40 years or older.

For a subset of biopsy specimens ($n = 5,705, 15.2\%$), testing results from the Linear Array HPV genotyping test (HPV LA; Roche Diagnostics, Indianapolis, IN) were available for the preceding cytology specimen.18 HPV genotyping of cytology specimens was done as a separate study, and specimens were selected independently of the biopsy specimen considered in this analysis. In this subset, the proportion of HPV16 was compared across laboratories. HPV16 was chosen because it is the most common and carcinogenic genotype.17 Because of the small sample size, the largest laboratory (L) was compared with all other laboratories combined for this subanalysis.

Analyses were performed using Stata 11.0 analytic software (StataCorp LP, College Station, TX). This study was approved by the University of New Mexico Human Research Review Committee.

Results

Across 15 laboratories, 37,486 biopsy specimens were processed and yielded an available result. Two hundred forty diagnoses (0.64%) wereequivocal and excluded from the analysis (CIN not otherwise specified [NOS], ASC-US, dysplasia, squamous intraepithelial lesion NOS, and lymphoma). The number of biopsy specimens ranged from 308 to 16,813 per laboratory (median, 794).

The main focus of the analysis was on the prevalence and reproducibility of CIN2; however, laboratories did not have similar patient populations. Principally, the severity of preceding cytology that led to the biopsy varied from 2.7% to 15.0% for HSIL or worse and from 2.6% to 10.2% for normal cytology ($P < .001$) Table 1. Patient age was also statistically different among laboratories ($P < .001$); mean age ranged from 28.0 to 33.2 years.

The crude percentage of diagnoses that were CIN2 varied from 7.1% to 22.3%, with 1 outlier at 1.5% (laboratory O) ($P < .001$) Table 2. Both severity of preceding cytology (AGUS/ASC-H/HSIL or worse) and age (40 or more years) were associated with an increased proportion of CIN2 diagnoses ($P < .001$ for both). When stratified by preceding cytology or age, the proportion of CIN2 diagnoses still varied significantly between laboratories ($P < .001$ for all strata). Because cytology was strongly associated...
Varied Cervical Histopathology Across 15 Laboratories

with histopathology result and variable across laboratories (Table 1), remaining calculations were directly standardized to the distribution of cytology results in laboratory L, the lab processing the most number of slides. Parallel analyses standardized to age yielded the same conclusions (results not shown). Simultaneous adjustment for cytology and age resulted in unstable estimates.

Figure 1 presents the overall distribution of biopsy specimen diagnoses across laboratories. We combined CIN2-3 diagnoses with CIN3 and cancer; this tended to reduce laboratory variability in the CIN2 vs CIN3 distinction (as compared with combining CIN2-3 with CIN2). Excluding laboratories with the largest and smallest values, the proportion of CIN2 diagnoses ranged from 6.2% to 14.4%
Anatomic Pathology / Original Article

Discussion

Our investigation of the statewide variability of cervical histopathology diagnoses showed a wide variation in the distribution of histologic results across laboratories. After adjusting for the preceding cytology and age of each case, the proportion of diagnoses that were CIN2 still varied considerably by laboratory. The amount of variability between laboratories, therefore, cannot be completely explained by differences in referral populations. More likely, this difference in the proportion of CIN2 diagnoses results from variability in histopathologist classifications between the diagnosis of CIN2 vs CIN1 and CIN2 vs CIN3. By comparison, the proportion of HPV16 associated with precancer tends to be a metric of disease severity in the population. If laboratories graded similarly, the proportion of HPV16 would be equal within each grade of CIN across laboratories. Yet, in a comparison of the largest laboratory with the others, the proportion of HPV16

(P < .001), varying less than unadjusted values presented in Table 2. Notably, the percentages of normal and CIN1 varied widely, with some laboratories calling more CIN1 vs normal diagnoses and vice versa (5.5%-57.7% for normal diagnoses and 23.3%-86.7% for CIN1 diagnoses; P < .001 for the CIN1:normal variability). Laboratories with a higher proportion of CIN1 diagnoses did not consistently yield a higher (or lower) proportion of CIN2 diagnoses.

Figure 1 Percent distribution of cervical biopsy specimen diagnoses by laboratory, standardized by preceding cytology and to laboratory L because it processed the greatest number of biopsy specimens. The figure excludes women with a cytology result if no diagnosis was given (n = 52) or if an unsatisfactory specimen was obtained (n = 45).

Consideration of the percentage of women diagnosed with HPV16 in the preceding cytology Table 3 showed a similar prevalence between laboratory L and the other laboratories (23.6% vs 24.0%; P = .691). Although the classification of CIN3 was associated with a similar proportion of HPV16 positivity (59.7% in laboratory L vs 61.1% in the other laboratories; P = .827), the proportion differed for CIN2 (39.0% in laboratory L vs 57.4% in the other laboratories; P < .001).
positivity in preceding cytology varied inversely with the prevalence of CIN2 diagnoses, suggesting systematic differences in classification by laboratory and/or pathologist with a preference for more sensitive vs stringent diagnosis.

We sought to control for the differing patient populations across laboratories by standardizing to the referral cytology. Because cytology itself is variable, we were unable to completely control for differences in underlying patient risk. We therefore used the more stringent metric of HPV16 positivity in the preceding cytology, which further substantiated our findings of variability between laboratories.

For 11% of the diagnoses, no preceding cytology result was available (defined as a cytology taken within 1 year of the biopsy specimen being processed). Of these diagnoses with missing cytology, 31.7% of biopsy specimens were processed in 2006, and their preceding cytology was likely unavailable because cytology results prior to January 2006 were not provided by all laboratories. It is also possible that specimens came from women undergoing a repeat colposcopy, and 43.9% of biopsy specimens missing cytology had a previous histology result of CIN2 or worse. Removing these biopsies in a sensitivity analysis did not change our findings. In addition, 5.8% of diagnoses had a normal previous cytology. These biopsy specimens were taken throughout the follow-up period and likely represent the reality of real-world clinical practice. These women might have had some abnormal cytology or histology result in the past that date prior to the initiation of the registry.

Recently, new guidelines were published by the College of American Pathologists and the American Society for Colposcopy and Cervical Pathology for the histologic classification of squamous intraepithelial lesions.18 They recommend using p16INK4a immunohistochemistry staining when considering a CIN2 diagnosis, with the goal to call the more definitively high- or low-grade CIN. In our study population, approximately 10% of diagnoses were CIN2 (virtually all of which would be p16INK4a tested) and four-tenths were CIN1, an unknown fraction of which would qualify for testing depending on the pathologist. Our findings suggest, therefore, that p16INK4a staining would be applied to a substantial minority of all biopsy specimens even if current CIN3 and normal diagnoses are not tested. As a caution regarding interpretation of our findings, the extent to which pathologists in this study used p16INK4a staining to manage equivocal lesions is not known, and it is possible that some of the differences in proportion of CIN2 diagnoses could be explained by differential adoption of p16INK4a staining in the laboratories.

Although not the focus of this analysis, we noted that some laboratories and, by inference, some pathologists systematically appeared to prefer CIN1 vs normal classifications and vice versa, suggesting that differentiating CIN1 from normal is particularly challenging. This is not a new observation. In support of this suggestion, data from the ASC-US/LSIL Triage Study demonstrated that the risk of subsequent CIN3 was similar between women with CIN1 and normal diagnoses at baseline.19

In conclusion, we documented large interlaboratory differences that must affect the treatment threshold. In our study, women would have a different probability of being treated depending on where their slides were sent. Although studies suggest that interobserver variability improves with utilization of p16INK4a staining, it remains unclear whether a de-emphasis of the CIN2 category and increased use of p16INK4a staining would improve interlaboratory variability and even accuracy. Implementation data are required to measure the impact of the new guidelines recommending systematic utilization of p16INK4a staining.

Table 3
Percentage of Women With an HPV16 Genotype Result at Previous Cytology,a Stratified by Laboratory L vs All Other Laboratories

<table>
<thead>
<tr>
<th></th>
<th>Laboratory L</th>
<th>Other Laboratories</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>Col %</td>
<td>% HPV16+</td>
</tr>
<tr>
<td>Total</td>
<td>2,756</td>
<td>100.0</td>
<td>23.6</td>
</tr>
<tr>
<td>Biopsy specimen result</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>1,332</td>
<td>48.3</td>
<td>16.7</td>
</tr>
<tr>
<td>CIN1</td>
<td>761</td>
<td>27.6</td>
<td>17.0</td>
</tr>
<tr>
<td>CIN2</td>
<td>421</td>
<td>15.3</td>
<td>39.0</td>
</tr>
<tr>
<td>CIN2-3</td>
<td>70</td>
<td>2.5</td>
<td>50.0</td>
</tr>
<tr>
<td>CIN3</td>
<td>149</td>
<td>5.4</td>
<td>59.7</td>
</tr>
<tr>
<td>Carcinoma</td>
<td>23</td>
<td>0.8</td>
<td>39.1</td>
</tr>
</tbody>
</table>

CIN, cervical intraepithelial neoplasia; HPV, human papillomavirus.

a Cytology taken within 1 year of biopsy specimen reading.

b Column percentages are not representative of the laboratory distribution because a stratified sample of specimens was selected. Biopsy specimen results between laboratory L and other laboratories were statistically different, P < .001.

From the 1Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health,
References


© American Society for Clinical Pathology

Am J Clin Pathol 2013;139:330-335
DOI: 10.1309/AJCPSD32XJX7P7NB