Should Atypical Squamous Cells of Undetermined Significance (ASCUS) Be Subcategorized?

Accuracy Analysis of Papanicolaou Smears Using Receiver Operating Characteristic Curves and Implications for the ASCUS/Squamous Intraepithelial Lesion Ratio

Andrew A. Renshaw, MD,1 David R. Genest, MD,2 and Edmund S. Cibas, MD2

Key Words: Pap smear; Cervicovaginal smear; Accuracy; Sensitivity; Diagnosis; Biopsy; Receiver operating characteristic curve

Abstract

We correlated all Papanicolaou test diagnoses over a 6-month period with biopsy results and determined accuracy using receiver operating characteristic curves and biopsy as the “gold standard.” Accuracies were calculated using all atypical squamous cells of undetermined significance (ASCUS) cases or by eliminating subsets thereof.

Retaining the ASCUS category resulted in significantly greater accuracy for the diagnosis of squamous intraepithelial lesion (SIL) on biopsy compared with eliminating it by diagnosing all such cases as negative. Subcategorization significantly improved the accuracy of the test only when all cases were included. The highest accuracy without subcategorization was achieved when ASCUS, favor reactive, cases were diagnosed as negative, but this threshold was significantly less sensitive than including all ASCUS cases. Increasing or decreasing the estimated ASCUS/SIL ratio from 2.4 without subcategorization significantly reduced accuracy. Similar results were obtained when high-grade SIL on biopsy was used as the gold standard.

Use of the ASCUS category significantly improves the accuracy of the Papanicolaou test. Eliminating any subset of ASCUS reduces the ASCUS/SIL ratio but also significantly diminishes the sensitivity of the Papanicolaou test.

The most commonly made abnormal diagnosis in the Bethesda System for reporting cervicovaginal cytology is atypical squamous cells of undetermined significance (ASCUS), constituting between 1% and 9% of Papanicolaou (Pap) test diagnoses (mean, 5%).1-3 Several studies have shown that subcategorizing ASCUS cases into favor reactive, not otherwise specified (NOS), favor squamous intraepithelial lesion (SIL), and favor high-grade SIL (HSIL) is associated with significantly different predictive values for a histologically confirmed SIL.4-7 Whether subcategorizing ASCUS improves diagnostic accuracy, however, is not known.

In addition, the ASCUS/SIL ratio has been promoted as a useful quality control measure.1,2 It has been suggested by the authors of the Bethesda System8 that the ASCUS/SIL ratio for an individual or laboratory should be less than 3.0, and others9 have suggested that lower ratios are desirable. Because the diagnosis of ASCUS conveys uncertainty, a low ratio reduces the uncertainty generated by the laboratory and may reduce the percentage of women with negative biopsy results. While this may be clinically desirable, it is not known whether a lower ASCUS/SIL ratio is associated with increased diagnostic accuracy.

To answer these questions, we correlated all diagnoses for Pap tests over a 6-month period with the results of subsequent tissue sampling (by biopsy or hysterectomy).

Materials and Methods

All Pap test diagnoses for a 6-month period from January 1 through June 30, 1995, from the Division of Cytopathology, Brigham & Women’s Hospital, Boston, MA, were reviewed...
and correlated with the results of corresponding tissue sampling (biopsy or hysterectomy) performed either concurrently or within 3 months of the smear being taken. This group of cases has been the subject of 2 earlier reports.4,10 Diagnoses were made in accordance with the Bethesda System.11 At the time of the original sign-out, ASCUS cases were qualified as favor reactive, favor SIL (no grade specified), or favor HSIL. Cases that were not qualified were categorized for this study as NOS.

The ASCUS/SIL ratio for the laboratory for the study period was 3.2. All ASCUS and LSIL rates and the ASCUS/SIL ratios reported herein are for the entire laboratory and are derived from the combination of all pathologists reviewing Pap smears during the study period. To estimate the resulting ASCUS/SIL ratio if subcategories of ASCUS were diagnosed as negative, it was assumed that the distribution of biopsies was proportional to the distribution of ASCUS diagnoses. So, for example, if half of the ASCUS cases with biopsies were now categorized as negative, it was assumed that half of the laboratory’s total ASCUS population would be categorized as negative.

Categorical analysis of proportions for sensitivity was done using a 2-tailed Pearson chi-square test.

Receiver operating characteristic (ROC) curves consist of graphs of sensitivity vs 1 minus specificity. The area under the curve (labeled A) represents the most popular overall diagnostic accuracy of the test and ranges from 0 to 1.0.12 An area of 0.5 represents an accuracy that would be achieved by chance alone, and an area of 1.0 is complete accuracy (100% sensitivity and specificity).

Results

The distribution of diagnoses on Pap test and the subsequent biopsy results are shown in Table 1.

The results of accuracy analysis with ROC curves as well as the sensitivity for SIL and the estimated ASCUS/SIL ratio using all SIL on biopsy as a gold standard are shown in Table 2.

By using SIL on biopsy as the gold standard, inclusion of the ASCUS category significantly increased the overall accuracy of the test compared with diagnosing all such cases as negative (A = 0.79 vs 0.73; P < .001). Subcategorization significantly improved the accuracy of the test when all cases were included (A = 0.81 vs 0.79; P < .001). Subcategorization

Table 1

<table>
<thead>
<tr>
<th>Cytologic Diagnosis</th>
<th>Negative Biopsy Result</th>
<th>SIL Biopsy Result</th>
<th>HSIL Biopsy Result Only</th>
<th>Total of Negative and SIL Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>304 (92.4)</td>
<td>25 (76)</td>
<td>13 (4.0)</td>
<td>329</td>
</tr>
<tr>
<td>ASCUS, favor reactive</td>
<td>93 (93.0)</td>
<td>7 (70)</td>
<td>4 (4.0)</td>
<td>100</td>
</tr>
<tr>
<td>ASCUS, NOS</td>
<td>151 (78.6)</td>
<td>41 (21.3)</td>
<td>18 (9.4)</td>
<td>192</td>
</tr>
<tr>
<td>ASCUS, favor SIL</td>
<td>89 (66.9)</td>
<td>44 (29.1)</td>
<td>23 (17.3)</td>
<td>133</td>
</tr>
<tr>
<td>ASCUS, favor HSIL</td>
<td>15 (56)</td>
<td>12 (44)</td>
<td>10 (37)</td>
<td>27</td>
</tr>
<tr>
<td>LSIL</td>
<td>92 (48.7)</td>
<td>97 (61.3)</td>
<td>40 (21.2)</td>
<td>189</td>
</tr>
<tr>
<td>HSIL</td>
<td>23 (18.4)</td>
<td>102 (81.6)</td>
<td>90 (72.0)</td>
<td>125</td>
</tr>
<tr>
<td>Total</td>
<td>767</td>
<td>328</td>
<td>198</td>
<td>1,095</td>
</tr>
</tbody>
</table>

ASCUS, atypical squamous cells of undetermined significance; HSIL, high-grade SIL; LSIL, low-grade SIL; NOS, not otherwise specified; SIL, squamous intraepithelial lesion.

* Data are given as number (percentage).

Table 2

<table>
<thead>
<tr>
<th>Categories Included</th>
<th>No. of ASCUS Categories</th>
<th>Area</th>
<th>Sensitivity</th>
<th>Estimated ASCUS/SIL Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>0</td>
<td>0.73 ± 0.02</td>
<td>0.34</td>
<td>0</td>
</tr>
<tr>
<td>All ASCUS cases</td>
<td>1</td>
<td>0.79 ± 0.02</td>
<td>0.96</td>
<td>3.2</td>
</tr>
<tr>
<td>Favor reactive, NOS, favor SIL, favor HSIL</td>
<td>4</td>
<td>0.81 ± 0.01</td>
<td>0.96</td>
<td>3.2</td>
</tr>
<tr>
<td>NOS, favor SIL, favor HSIL</td>
<td>1</td>
<td>0.81 ± 0.02</td>
<td>0.78</td>
<td>2.4</td>
</tr>
<tr>
<td>NOS, favor SIL, favor HSIL</td>
<td>3</td>
<td>0.81 ± 0.02</td>
<td>0.78</td>
<td>2.4</td>
</tr>
<tr>
<td>Favor SIL, favor HSIL</td>
<td>1</td>
<td>0.79 ± 0.02</td>
<td>0.51</td>
<td>1.1</td>
</tr>
<tr>
<td>Favor SIL, favor HSIL</td>
<td>2</td>
<td>0.79 ± 0.02</td>
<td>0.51</td>
<td>1.1</td>
</tr>
<tr>
<td>Favor HSIL</td>
<td>1</td>
<td>0.79 ± 0.02</td>
<td>0.36</td>
<td>0.5</td>
</tr>
</tbody>
</table>

ASCUS, atypical squamous cells of undetermined significance; HSIL, high-grade SIL; NOS, not otherwise specified; SIL, squamous intraepithelial lesion.

* Any biopsy-proven SIL considered a positive result. An area of 0.5 represents an accuracy that would be achieved by chance alone, and an area of 1.0 is complete accuracy (100% sensitivity and specificity).
did not increase diagnostic accuracy at any other threshold. The highest accuracy without subcategorization was achieved when all cases diagnosed as ASCUS, favor reactive, were classified as negative (A = 0.81). This was associated with an estimated ASCUS/SIL ratio of 2.4, but this threshold was significantly less sensitive (0.76 vs 0.96; \( P < .001 \)) than including all ASCUS cases (ASCUS/SIL ratio, 3.2). Further decreasing the estimated ASCUS/SIL ratio by progressively raising the threshold for ASCUS (with or without subcategorization) significantly reduced accuracy (\( P < .001 \)).

A similar analysis using only a diagnosis of HSIL as the gold standard on biopsy is given in Table 3. The results were essentially the same. Without subcategorizing, the most accurate result was obtained when cases originally classified as ASCUS, favor reactive, were reclassified as negative (A = 0.81). However, the same accuracy could be obtained by including these cases and subclassifying them.

The specific value of individual subcategories is explored further in Table 4. In this analysis, all ASCUS cases were always diagnosed as ASCUS, but specific subcategories were segregated differently. HSIL was used as the gold standard. In this approach, no significant increase in overall accuracy was obtained when individual categories such as ASCUS, favor HSIL, were subcategorized.

**Discussion**

The results we present herein suggest that cytopathologists, when not allowed to subcategorize ASCUS cases, must make a choice between accuracy and sensitivity. Diagnosing all cases that would otherwise be classified as ASCUS, favor reactive, as negative results in the highest accuracy. However, this results in significantly lower sensitivity. If one is willing to subcategorize, then the same accuracy can be achieved with a significantly higher sensitivity at the price of an increased number of ASCUS cases and a higher ASCUS/SIL ratio. When interpreting smears, cytologists must make a choice between sensitivity and ASCUS rate. This choice applies whether one chooses all SIL or only HSIL as the gold standard. Which approach is best? The answer to this must be placed in clinical context. We believe that because the Pap test is a screening test, its clinical value is very dependent on its sensitivity. Eliminating the ASCUS, favor reactive, category significantly reduces sensitivity and, when subcategorization is allowed, does not decrease diagnostic accuracy.

But our data suggest that this trade-off between sensitivity and accuracy works only within a limited range. Reducing the ASCUS/SIL ratio further by raising the lowest threshold for the diagnosis of ASCUS further (by diagnosing all ASCUS,
NOS, cases as negative) actually results in decreased accuracy. Although one may achieve a lower ASCUS/SIL ratio by lowering the threshold for SIL (ie, by categorizing cases such as ASCUS, favor SIL, and ASCUS, favor HSIL, as LSIL and HSIL, respectively), further reductions at the lower threshold decrease not only sensitivity but also accuracy. Of course, the results presented herein are specific to this laboratory setting, and the exact threshold may vary from laboratory to laboratory. In addition, we do not have an ASCUS/SIL population with a ratio over 3.2, so it is not clear whether increasing the ASCUS/SIL ratio above 3.2 and increasing the number of cases diagnosed as ASCUS above what was diagnosed in the present study would be associated with improved or decreased test performance. Nevertheless, these data strongly suggest that there is an optimum ASCUS/SIL ratio and that increasingly lower ratios are not necessarily associated with increasing diagnostic accuracy. While the ASCUS/SIL ratio has provided a reasonable surrogate marker to evaluate the appropriate use of the ASCUS diagnosis, the results of the present analysis seem to offer a more direct, meaningful, and useful assessment of whether the diagnosis of ASCUS (and the associated ASCUS/SIL ratio) is being used appropriately within a laboratory.

Second, while subcategorization improves the diagnostic accuracy of the Pap smear when cases diagnosed as ASCUS, favor reactive, are included, it is very difficult to show that subcategorization outside this setting is of value. In no other setting does subcategorization add value. This is somewhat counterintuitive, since the subcategory with the strongest correlation with biopsy findings is ASCUS, favor HSIL. However, the number of cases that qualify for this diagnosis is so low that it is insufficient to raise the overall accuracy of the test, regardless of whether SIL or HSIL is used as the gold standard. Since this category may have clinical importance, however, in that some physicians may recommend colposcopy and biopsy rather than a repeated Pap test for patients with this diagnosis, it is clinically justified to subcategorize these cases even though they do not significantly affect the overall performance of the test.

The method we present in this article seems to be a very useful way to evaluate the performance of Pap test interpretation in relation to the results of follow-up biopsy. Our results show that retaining the ASCUS category significantly improves the accuracy of the Pap smear compared with diagnosing all such cases as negative. Subcategorization significantly improves the accuracy of the Pap test only when all cases, including ASCUS, favor reactive, are included. The most accurate threshold for ASCUS without subcategorization is obtained when all ASCUS, favor reactive, cases are diagnosed as negative, which has an estimated ASCUS/SIL ratio of 2.4 but is significantly less sensitive and no more accurate than when all ASCUS cases are diagnosed and subcategorized.

From the Departments of Pathology, Baptist Hospital of Miami, Miami, FL, and the Brigham & Women’s Hospital and Harvard Medical School, Boston, MA.

Acknowledgment: We thank Kelly H. Sou, PhD, Department of Radiology, Brigham & Women’s Hospital, and Department of Health Care Policy, Harvard Medical School, Boston, MA, for review of the statistical methods and results.

References