

Subcategorization of Papanicolaou Tests Diagnosed as Atypical Squamous Cells of Undetermined Significance

Stephen S. Raab, MD

In this issue of the *Journal*, Renshaw et al¹ address the issue of subcategorizing Papanicolaou (Pap) tests diagnosed as atypical squamous cells of undetermined significance (ASCUS). I would like to make several comments about this article in light of the recommendations presented on the Web site of the 2001 Bethesda System Conference² and in light of ancillary studies, such as human papillomavirus (HPV) testing.

One of the recommendations from the 2001 Bethesda System Conference is for changing the terminology of Pap tests formerly called ASCUS (using the 1991 Bethesda System terminology).^{2,3} The 1991 Bethesda Committee recommended that ASCUS Pap tests could be subclassified as favor reactive or favor dysplasia (favor low-grade squamous intraepithelial lesion [SIL]), and if neither of these categories seemed to fit, then the Pap test could be diagnosed simply as ASCUS (which some prefer to term the not otherwise specified [NOS] category).³ The 2001 Bethesda Committee recommended eschewing previous ASCUS terms for the term atypical squamous cells (ASC); all Pap tests then should be subclassified as either atypical squamous cells–undetermined significance (ASC-US) or atypical squamous cells–high-grade dysplasia not excluded (ASC-H).²

Note the differences in the 1991 and the 2001 Bethesda Systems categories of ASCUS. It was argued that the ASCUS, favor reactive category be eliminated from the 2001 schema because follow-up reveals that the majority of women with this diagnosis do not have a high-grade dysplasia but have a benign lesion.² The ASCUS, NOS and ASCUS, favor dysplasia categories were collapsed into 1 category because it was thought that both categories included women with a significant risk of having dysplasia (mostly

low grade) on follow-up. Although the risk of dysplasia associated with the categories of ASCUS, NOS and ASCUS, favor dysplasia are different, this difference in risk was not considered disparate enough to warrant 2 categories. Follow-up data for women with ASC-H show that the risk of having dysplasia, and particularly high-grade dysplasia, is significantly higher than for women with ASC-US.² Atypical cells classified as ASC-H usually are metaplastic, and it is interesting to note that the 1991 Bethesda Committee recognized these cells as potentially representing a high-grade dysplasia.³ As a result of these recommendations, the 1991 Bethesda System category of ASCUS morphed from 3 categories to 2 categories, one predominantly a risk category for low-grade dysplasia and the other predominantly a risk category for high-grade dysplasia.

The ASCUS categories examined by Renshaw et al¹ are a hybrid of the 1991 and the 2001 Bethesda categories. The 1991 ASCUS, favor reactive category is maintained; the 1991 ASCUS, favor dysplasia and ASCUS, NOS categories are not collapsed; and the 2001 ASC-H category is added. Using receiver operating characteristic curve analysis, Renshaw et al¹ argued that the sensitivity of the Pap test is lowered if any ASCUS subset is eliminated. Does this mean that the 2001 Bethesda Committee is mistaken in eliminating ASCUS, favor reactive and in collapsing ASCUS, NOS and ASCUS, favor dysplasia? I believe that the 2001 Bethesda Committee is justified in its recommendations, although I agree with Renshaw et al¹ that removal of ASCUS subcategories decreases Pap test sensitivity.

Before discussing these points further, the data and methods of Renshaw et al¹ first deserve some comment. The data used in their analysis suffer from sample bias. The

Table 1
Histologic Follow-up in an Imaginary Laboratory With a Sensitivity of 50% and a Specificity of 98%

Papanicolaou Test Diagnosis	Histologic Follow-up	
	Negative	Cervical Intraepithelial Neoplasia
No evidence of squamous intraepithelial lesion or malignancy	9,200	300
Atypical squamous cells—undetermined significance	150	150
Squamous intraepithelial lesion	50	150
Total	9,400	600

population studied was predominantly a colposcopic population, and the selection of patients for disease verification based on tissue biopsy results in both a high prevalence of histologic abnormalities and “workup” bias.^{4,5} This latter bias results in lowered estimates of specificity and elevated estimates of sensitivity.⁶ Renshaw et al¹ reported that when ASCUS diagnoses were included, the Pap test sensitivity was 96%; although previous and unpublished work must be analyzed to determine its reported Pap test specificity, it would seem from the data of Renshaw et al¹ that the specificity was less than 50%. In meta-analyses of Pap test performance measures, authors have reported that the sensitivity of the Pap test is in the range of 50% and the specificity is in the range of 98% (apparently the opposite of the data reported by Renshaw et al).^{4,7}

In their calculation of the accuracy of the Pap test, Renshaw et al¹ assumed that if Pap tests were not diagnosed using an ASCUS subcategory, they would be diagnosed as negative. In actual practice, this may not be a valid assumption because if given the choice of using or not using an ASCUS subcategory, it is uncertain whether pathologists would upgrade or downgrade these Pap tests. Few data exist on this subject. At the 2000 Papanicolaou Society Companion Meeting at the United States and Canadian Academy of Pathology Annual Meeting, Frable presented data showing that after discontinuation of the entire ASCUS category in his laboratory, the SIL percentage rose considerably.⁸ This would indicate that most ASCUS Pap tests were upgraded to SIL (the opposite of what Renshaw et al assume); these data have yet to be published in a peer-reviewed journal. The point I am trying to make is that it is uncertain what diagnoses would be used if ASCUS were eliminated. Consequently, the assumption that if the ASCUS category (or any of its subcategories) is eliminated then Pap tests formerly diagnosed as ASCUS would now be diagnosed as benign needs justification. The elimination of ASCUS was discussed at the 2001 Bethesda Conference and rejected.²

Receiver operating characteristic curve analyses have been used to examine the accuracy of diagnostic thresholds (eg, ASCUS/cervical intraepithelial neoplasia [CIN] 1 vs

low-grade SIL/CIN 1) across articles reporting Pap test accuracy.⁴ In measuring the accuracy of any test, a tradeoff is made between sensitivity and specificity as a particular threshold is chosen. McCrory et al⁴ reported summary “effectiveness” scores for different thresholds, representing the tradeoff between sensitivity and specificity. For example, the effectiveness score for ASCUS/low-grade SIL was 1.027 (95% confidence interval, 0.777-1.144) and for high-grade SIL/CIN 2-3 was 1.287 (95% confidence interval, 1.075-1.499).⁴ The higher effectiveness threshold for high-grade lesions indicates that the Pap test has a better combined sensitivity and specificity for detecting lesions at the high-grade threshold than it does at the ASCUS/CIN 1 threshold.

Neither the data of McCrory et al⁴ nor the bias in the study by Renshaw et al¹ necessarily refute the conclusions made by Renshaw et al. If Pap tests formerly diagnosed as a subcategory of ASCUS now are classified as negative, the sensitivity of the Pap test still decreases. However, Renshaw et al¹ did not emphasize that as the sensitivity of the Pap test decreased, the specificity of the Pap test increased. Imagine that a laboratory has a 5% epithelial cell abnormality rate and an ASCUS/SIL ratio of 3:2, and follow-up of women with ASCUS Pap tests shows that 50% have dysplasia. Also imagine that the Pap test is 50% sensitive and 98% specific. **Table 1**. If the ASCUS Pap tests were reclassified as benign, the resulting sensitivity and specificity would be 25% and 99%, respectively. The overall accuracy of the Pap test would also decrease (as Renshaw et al indicated). In my mind, the elimination of the entire ASCUS category would not be acceptable because the loss in sensitivity would not be justified by the gain in specificity, which is a different point than what is argued by Renshaw et al.¹ A loss in sensitivity and gain in specificity would also occur if Pap tests classified in some of the ASCUS subcategories were reclassified as benign. However, if a pathologist could accurately reclassify a percentage of ASCUS Pap tests, the resulting sensitivity-specificity tradeoff may be justified.

As stated, the 2001 Bethesda Committee recommended discontinuing the ASCUS, favor reactive category.² Behind this recommendation was the belief that the majority of these

Pap tests would be diagnosed as negative and a small percentage would be diagnosed as ASC-US. How pathologists would actually do this was not stated, and it is unknown whether some pathologists would simply diagnose these Pap tests as ASC-US. Thus, the jury is still out about whether the elimination of this ASCUS subcategory would result in decreased Pap test accuracy.

I think the decision to discontinue or add categories is partly one of managing risk, setting thresholds, and assessing clinical import. This may be illustrated by examining the benign cellular changes (BCC) category, which was eliminated by the 2001 Bethesda Committee.² The 1991 Bethesda System category of BCC is a higher risk category for dysplasia (mostly low-grade) compared with the category of within normal limits (WNL).^{3,9,10} However, the clinical significance of this difference in risk is uncertain. A good classification system should have relatively specific clinical management strategies that correspond to the specific diagnostic categories. Retaining the category of BCC increases the accuracy of the Pap test, but this increased accuracy has no current clinical import because most expert clinicians view this diagnostic category to be of low risk (particularly for high-grade dysplasia) and recommend that women with BCC should not be treated differently from women with a diagnosis of WNL.² Most expert gynecologists think that women with BCC should have a follow-up Pap test in 12 months and not be followed up more aggressively.² The problem of eliminating BCC is that in the future, a specific management arm may become appropriate for this patient group. Alternatively, an ancillary test (such as the HPV test¹¹) could be used to triage this patient group into those of high and those of low risk. In summary, the improved accuracy of the Pap test associated with using BCC is important if specific management strategies or ancillary tests are available (or could become available). I believe that the BCC category was eliminated partly because clinicians found this category confusing. If both WNL and BCC are considered “benign” diagnoses, the argument is that they should be grouped together (despite the loss in Pap test sensitivity) and, hence, the birth of the 2001 Bethesda System category of negative for intraepithelial lesion or malignancy (NIL).²

I think the rationale of combining the ASCUS, NOS and ASCUS, favor dysplasia categories also was based partly on clinical reasons. Although the continuation of these 2 separate ASCUS categories may improve Pap test accuracy (as shown by Renshaw et al¹), most expert clinicians recommend that the treatment should be the same for women who have either diagnosis.² Consequently, I believe that these 2 ASCUS categories were combined for simplicity; the improved accuracy of the Pap test by maintaining these 2 categories was irrelevant. Similarly, if

pathologists wanted, they probably could reclassify low-grade SIL Pap tests into 2 subcategories of differing risk (eg, low-grade SIL, less likely to be dysplasia and low-grade SIL, more likely to be dysplasia), which would improve Pap test accuracy; again, the clinical import of these 2 categories would be questionable.

Guidelines for treating women with specific Bethesda System diagnoses will follow the American Society for Colposcopy and Cervical Pathology Consensus Conference that was held September 6-9, 2001. However, based on the Atypical Squamous Cells of Undetermined Significance/Low-Grade Squamous Intraepithelial Lesions Triage Study Group data and the 2001 Bethesda Conference, it seems apparent that one of the possible treatments for women with ASC-US is HPV testing for triage.^{2,11} HPV testing probably will result in changes in ASCUS follow-up data and in how pathologists use ASC-US. Laboratories may report HPV testing data in different ways.² Some laboratories may use the HPV test result in an interpretive manner and thereby reinterpret Pap tests diagnosed as ASC-US into the category of low-grade SIL or the category of NIL; if this method of reporting is adopted, the ASC-US category will be used less frequently. Other laboratories may interpret the HPV result probabilistically and simply report the HPV test result along with the ASC-US diagnosis and a probabilistic statement; in this reporting method, the ASC-US category remains. A potential consequence of HPV testing is the overuse of the ASC-US category. Instead of deciding whether a Pap test is NIL or ASC-US or SIL, pathologists may be more likely to make a diagnosis of ASC-US and use HPV testing for triage. Depending on reimbursement levels, a financial bias could result in the overinterpretation of ASC-US cases. The consequences of HPV testing on Pap test accuracy are unknown, and methods such as those used by Renshaw et al¹ will be helpful for this assessment.

From the Department of Pathology and Laboratory Medicine, Allegheny General Hospital, Pittsburgh, PA.

References

1. Renshaw AA, Genest DR, Cibas ES. 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. *Am J Clin Pathol.* 2001;116:692-695.
2. National Cancer Institute Bethesda System 2001. Web site available at: www.Bethesda2001.cancer.gov. Accessed September 8, 2001.
3. Kurman RJ, Solomon D. *The Bethesda System for Reporting Cervical/Vaginal Cytologic Diagnoses*. New York, NY: Springer-Verlag; 1994.

4. McCrory DC, Mather DB, Bastain L, et al. *Evaluation of Cervical Cytology*. Rockville, MD: Agency for Health Care Policy and Research; February 1999. Evidence Report/Technology Assessment No. 5. AHCPR publication 99-E010.
5. Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med*. 1978;299:926-930.
6. Feinstein AR. *Clinical Epidemiology: The Architecture of Clinical Research*. Philadelphia, PA: Saunders; 1985:434.
7. Fahey MT, Irwig L, Macaskill P. Meta-analysis of Pap test accuracy. *Am J Epidemiol*. 1995;141:680-689.
8. Frable J. A rationale for the elimination of the use of ASCUS. Paper presented at the Papanicolaou Society Companion Meeting, United States and Canadian Academy of Pathology Annual Meeting; March 25, 2000; New Orleans, LA.
9. Barr Soofer S, Sidawy MK. Reactive cellular change: is there an increased risk for squamous intraepithelial lesions? *Cancer*. 1997;81:144-147.
10. Malik SN, Wilkinson EJ, Drew PA, et al. Benign cellular changes in Pap smears: causes and significance. *Acta Cytol*. 2001;45:5-8.
11. The Atypical Squamous Cells of Undetermined Significance/Low-Grade Squamous Intraepithelial Lesions Triage Study (ALTS) Group. Human papillomavirus testing for triage of women with cytologic evidence of low-grade squamous intraepithelial lesions: baseline data from a randomized trial. *J Natl Cancer Inst*. 2000;92:397-402.