ASC, TBS, and the Power of ALTS

Mark H. Stoler, MD

DOI: 10.1309/J31A0YQM37Q4V4KK

In this issue of the Journal, Lee and colleagues from Hong Kong describe their assessment of the impact of the 2001 Bethesda System (TBS) revisions on their cytology practice. The primary focus of the study was to assess the impact of the conversion from TBS 1991 to the 2001 version on the frequency of atypical squamous cells (ASC) diagnoses. As noted in prior commentaries, ASC is not a diagnosis. ASC is a poorly reproducible mixture of reactive and inflammatory changes that mimic squamous intraepithelial lesion (SIL) in a group of patients who have SIL, mostly low grade (ASC-US [undetermined significance]) but perhaps 5% to 10% high grade (ASC-H). So, can changes in classification terminology and criteria impact how much we equivocate? Let’s see.

Cases of ASC were retrospectively correlated within their computerized database for higher-grade end points including low-grade SIL (LSIL)/cervical intraepithelial neoplasia (CIN) 1, HSIL/CIN 2 or 3, or carcinoma based on routine practice. Fourteen cytopathologists and an undetermined number of cytotechnologists were involved in this study, which basically compared 28,000 cases from July 2000 to June 2002 when TBS 1991 was used with a similar number of cases in the 2-year period from July 2002 to June 2004. The major conclusion in this study was that the conversion to TBS 2001 led to a small but statistically significant increase in the specificity of the ASC diagnostic category and virtually all other diagnostic rates were unchanged. Is this really an important finding? And does it support the conversion to TBS 2001 as an important advance, as the authors suggest?

Unfortunately, the comparisons in this article are confounded by multiple variables. The terminology conversion undoubtedly had a learning curve such that the application of the revised criteria early in the second observation period may have yielded different rates compared with when the operators were more familiar with the criteria later in the study. Regrettably, few time trend data are included. It is important to note that the terminology conversion was also associated with the implementation and partial conversion from conventional spray-fixed Papanicolaou smears to liquid-based cytology preparations. In the detail of the report it is also noted that the collection devices used for conventional cytology included only the spatula and no endocervical sampling device compared with the broom device used for liquid-based cytology preparations. In addition, the period of study was associated with a region-wide campaign to increase awareness of the importance of cervical cancer screening, which the authors suggest brought many more high-risk patients into the screening system. Follow-up was not systematic and was limited to the routine practice of 10 gynecologists. The relative proportion of smears and the types of preparations per gynecologist were not controlled for, and liquid-based cytology, which seemed to perform somewhat better on some measures, was biased in its use toward high-risk patients. Thus, any interpretation of the impact of the terminology and criteria change brought about by TBS 2001 is confounded by these many factors, some of which would favor some of the observed outcomes and others of which would oppose or tend to minimize the potential differences between the 2 observations.

Of course, increased awareness of the importance of cervical cancer screening, the conversion to liquid-based cytology, and the changes in terminology are interacting factors that apply not only to our colleagues in Hong Kong but also to many other countries heavily involved in cervical cancer screening, including the United States. It is remarkable how similar some of the findings reported by Lee and colleagues1...
are to those reported by the College of American Pathologists in their 2003 review of the impact of TBS 2001 on reporting rates compared with prior periods. Lee and colleagues do a good job in making these comparisons and rationalizing the similarities and differences between countries.

Even more remarkable, there is virtually no mention of human papillomavirus (HPV) testing and its usefulness in adjudicating the accuracy of cytologic diagnoses in this report. Lee and et al noted a decrease in LSIL/CIN 1 rates with age, which they think is related to the decreasing frequency of HPV infection with age, a well-recognized phenomenon in epidemiologic data sets. However, the fact remains that descriptive statistical studies such as those reported here and by the College of American Pathologists, although useful for benchmarking, do not get at the more critical issue of test performance. It is one thing to say that a laboratory has an ASC-US rate of 5% or 10% and ASC/SIL ratio of 1.5 or 2.0. It is quite another thing to say that 50% of the cases called ASC-US are actually HPV+ compared with 80% or 90% of the SIL calls and 10% of the negative cases. The former analysis suggests that different groups are reading slides roughly the same. The other analysis documents that the different groups are actually detecting the relevant biology at roughly the same rates. Thus, in my opinion, although certain statistical and numeric data are required by law to be tabulated and followed up in the United States, the implementation of HPV testing allows for more accurate and insightful analysis of the potential accuracy of diagnostic interpretations on cytologic specimens.

In the absence of HPV testing, can we say anything more insightful about the performance of our colleagues from Hong Kong as they transition into a screening and reporting system that is in many ways very similar to that in the United States? Although the follow-up was opportunistic and the assessment of events confounded by the limitations of the follow-up system, it is interesting to observe that the rate of HSIL/CIN 2 or 3 in this population was reported at approximately 10% during the first 6 to 12 months of follow-up. There were no statistically significant differences between the rates of high-grade end points among the different terminology systems or cytologic preparations, yet the rate of 8% to 10% is quite grade end points among the different terminology systems or statistically significant differences between the rates of high-

<table>
<thead>
<tr>
<th>Years of Cytology/HPV Stratified by Age*</th>
<th>ASC-US</th>
<th>ASC-US/HPV+</th>
<th>ASC-H</th>
<th>HSIL (ALTs)</th>
<th>HSIL (outright)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>18-25</td>
<td>15-20</td>
<td>25-30</td>
<td>45-50</td>
<td>65-70</td>
<td>80-85</td>
</tr>
<tr>
<td>36-45</td>
<td>5-10</td>
<td>20-25</td>
<td>30-40</td>
<td>60-70</td>
<td>85-90</td>
</tr>
<tr>
<td>≥46</td>
<td>10-20</td>
<td>20-25</td>
<td>40-50</td>
<td>85-90</td>
<td>80-90</td>
</tr>
</tbody>
</table>

*Age stratified estimates are based on actual ALTS data using quality control group readings of the enrollment liquid-based cytology preparations rounded off and provided by Phillip Castle, National Cancer Institute, Bethesda, MD (personal communication). Data are given as percentages.

© American Society for Clinical Pathology
HPV+ fraction or CIN follow-up data that fall significantly below the ALTS benchmarks, there are 2 possible choices. One is that there is a systematic difference in cytologic interpretation compared with the rates reported in ALTS. The other, more troublesome, is when laboratory professionals infer, for example, that their population is at lower or higher risk or infected by a different spectrum of HPV types. “My lab is different,” they declare; therefore, the benchmarks are wrong and do not apply. Put together a bunch of laboratories with this attitude, and regression to the mean leads to a change in the group performance standards: “It is OK that only 30% of my ASC diagnoses are HPV+ because I don’t overcall ASC; my patients are different.”

The careful statistical design of ALTS and other similar trials, which controlled for numerous variables, makes this defensive interpretation erroneous. The research community has given considerable effort to the development of a system of benchmarks that are reliable and representative of most populations in the United States, and, given the relatively universal similarity of HPV type distributions, particularly HPV-16, among populations worldwide, these benchmarks are, I believe, reasonably generally applicable.13 It would not be such a terrible thing for laboratories to calibrate their interpretations to these established literature benchmarks through a process of education and feedback. Ultimately, such an effort could lead to near-universal and reliable predictive values for cytologic interpretation, which directly impacts how our clinical colleagues follow up and treat patients.

From the Robert E. Fechner Laboratory of Surgical Pathology, University of Virginia Health System, Charlottesville.

References