

# Results of Multiple-Slide, Blinded Review of Papanicolaou Slides in the Context of Litigation

## *Determining What can be Detected Regularly and Reliably*

Andrew A. Renshaw, M.D.<sup>1</sup>  
E. Blair Holladay, Ph.D.<sup>2</sup>  
Karen B. Geils, M.S.<sup>2</sup>

<sup>1</sup> Department of Pathology, Baptist Hospital of Miami, Miami, Florida.

<sup>2</sup> Center for Quality Improvement in Gynecologic Cytopathology; Medical University of South Carolina, Charleston, South Carolina.

**BACKGROUND.** Multiple-slide, blinded review has been endorsed by several cytology and pathology organizations as the most appropriate method for the review of cervicovaginal specimens in the context of litigation. This process involves review of litigation slides in a blinded manner by multiple independent cytotechnologists and the comparison of those results with those of validation cases that are comprised of known abnormalities with biopsy follow-up. To the authors' knowledge, the results of this method have not been previously published.

**METHODS.** The results of the blinded review program at the Center for Cytopathology and Molecular Research at the Medical University of South Carolina for the years 1998–2004 were reviewed.

**RESULTS.** A total of 135 litigation slides and 122 validation slides were reviewed. The interpretations of these cases were found to be significantly different ( $P < 0.001$ ). Litigation cases were significantly more likely to be interpreted as either negative for intraepithelial lesion (NIL) or atypical squamous cells/atypical glandular cells (ASC/AGC) ( $P < 0.001$ ). The results appeared to be independent of the individual cytotechnologists involved. Approximately 10% of litigation cases were called at least ASC/AGC by all observers and 4% were interpreted as NIL by all observers. For litigation cases, ASC/AGC was found to be just as reproducible as high-grade squamous intraepithelial lesion. The results demonstrated that only 10% of litigation cases are regularly and reliably identified as abnormal, whereas a single review as performed by an expert cytologist can be expected to classify 56% of cases as abnormal.

**CONCLUSIONS.** This program suggests that a majority of litigation cases are not regularly and reliably identified as abnormal, and a single review will routinely overestimate the percentage of cases that are identified regularly and reliably. *Cancer (Cancer Cytopathol)* 2005;105:263–9. © 2005 American Cancer Society.

**KEYWORDS:** pathology, cytopathology, accuracy, litigation, blinded review, sensitivity, quality.

**M**ultiple-slide, blinded review of Papanicolaou (Pap) slides in the context of litigation has become the standard litmus test, and multiple national and regional cytopathology societies have endorsed this as the most appropriate method with which to distinguish acceptable errors (i.e., errors that could occur in any laboratory) from unacceptable errors that fall below the standard of care.<sup>1–9</sup> To our knowledge to date, little information concerning the results of this method are available. At the Medical University of South Carolina, this method has been used since 1995. We sought to review the results of multiple-slide, blinded review of litigation cases in the context of

Address for reprints: Andrew A. Renshaw, M.D., Department of Pathology, Baptist Hospital of Miami, 8900 N. Kendall Dr., Miami, FL 33176; Fax (786) 596-5986; E-mail: AndrewR@bhssf.org

Received March 11, 2005; revision received May 20, 2005; accepted May 26, 2005.

litigation to assess the value of this method and suggest methods for improvement.

## MATERIALS AND METHODS

The results of a blinded review in the context of litigation at the Medical University of South Carolina from 1998–2004 were reviewed. The details and specific methodologies of this program have been reviewed in previous publications.<sup>10</sup> In brief, all identifiers/markers are removed from the litigation slides. The slides are placed in a group of routine cases and validation slides for a total of 60 slides. Each group of 60 cases has only 1 litigation slide. Validation slides are comprised of abnormal cases with known biopsy follow-up that the reviewers are expected to identify in the process of screening the 60 cases. As is the usual practice in the clinical setting, all cases are reviewed by cytotechnologists; cytopathologists were not involved in screening of these cases. If the reviewer fails to identify the validation cases at the appropriate rate, then the results of that reviewer are not used to evaluate the litigation slides. In general, cases are reviewed 10 times by 10 separate observers, although in the early part of the program select cases were only reviewed as few as 6 times. Cytotechnologists' diagnoses were made using the Bethesda system in use at that time. For purposes of this study, these interpretations have been updated using the most recent Bethesda guidelines, Bethesda 2001.<sup>11</sup> For purposes of analysis, interpretations of atypical squamous cells (ASC) and atypical glandular cells (AGC) were grouped together (ACS/AGC), and carcinoma cases included both squamous cell carcinoma and adenocarcinoma.

The goal is to identify and distinguish cases that can be regularly and reliably identified from those that cannot. "Identifying a case" means determining that it is abnormal. In general, in the context of this process, it is rendering an interpretation that is anything other than negative for intraepithelial lesion (NIL), although the specific interpretations that might be included in this process can be varied and are specifically varied in the results section. Formal definitions of what results correspond with cases that are regularly and reliably identified have been defined previously.<sup>12</sup> In general, those cases that are regularly and reliably identified are cases that one would expect the average laboratory to be able to identify.

Ease of interpretation is a measure of how often a case can be identified based on repeated screenings. As all cytologists know, some cases are easy to identify as abnormal and others are very difficult. Ease of interpretation is simply a more formal quantitative measure of how difficult an individual slide might be to identify as abnormal. Ease of interpretation is the

**TABLE 1**  
Distribution of Interpretations for Litigation Slides (*n* = 135)

	NIL	ASC/AGC	Unsatisfactory	LSIL	HSIL	CA	Total
No.	61	575	335	77	180	69	1305
%	5	44	26	6	14	5	

NIL: negative for intraepithelial lesion; ASC/AGC: atypical squamous cells/atypical glandular cells; LSIL: low-grade squamous intraepithelial lesion; HSIL: high-grade squamous intraepithelial lesion; CA: carcinoma.

**TABLE 2**  
Distribution of Interpretations for Validation Slides (*n* = 122)

	NIL	ASC/AGC	Unsatisfactory	LSIL	HSIL	CA	Total
No.	4	24	83	267	562	249	1189
%	0.3	2	7	22	47	21	

NIL: negative for intraepithelial lesion; ASC/AGC: atypical squamous cells/atypical glandular cells; LSIL: low-grade squamous intraepithelial lesion; HSIL: high-grade squamous intraepithelial lesion; CA: carcinoma.

opposite of how difficult or hard a case might be to identify as abnormal. Cases that are always identified as abnormal (i.e., those that are identified as abnormal 100% of the time) have an ease of interpretation of 100%. Cases that are often but not always identified (i.e., those identified 80% of the time) have an ease of interpretation of 80%, and so forth. An ease of interpretation can be assigned to an individual slide or to a group of slides. To measure the ease of interpretation of a single slide, one simply needs to screen it multiple times and the ease of interpretation will be the percentage of the time that it is determined to be abnormal. For a group of slides, one must rescreen the group of slides; the ease of interpretation is the percentage of the time the group of slides is determined to be abnormal (i.e., the average). As an average, it is possible that individual slides within the group may have very different eases of interpretation compared with the group as a whole.

Statistical analysis was performed using a two-tailed Fisher exact test, a two-tailed chi-square test, and a nonparametric Kruskal–Wallis test as appropriate. A *P* value < 0.05 was considered statistically significant.

## RESULTS

Cytotechnologists' interpretations of 135 litigation slides and 122 validation slides were reviewed. The distribution of interpretations for the litigation slides and validation slides are shown in Tables 1 and 2. The distribution of interpretations for the litigation slides was found to be significantly different from that for

**TABLE 3**  
**Number and Percentage of Slides Regularly and Reliably Detected Using Different Thresholds for Litigation Slides (n = 135)**

No. of interpretations	A	B	C	D	E
10 of 10	0	2 (1)	2 (1)	2 (1)	14 (10)
9 of 10	0	4 (3)	6 (4)	8 (6)	33 (24)
8 of 10	1 (1)	6 (4)	11 (8)	13 (10)	43 (32)
7 of 10	1 (1)	9 (7)	15 (11)	17 (13)	52 (39)
6 of 10	1 (1)	12 (9)	22 (16)	26 (19)	65 (48)
5 of 10	1 (1)	21 (16)	29 (21)	34 (25)	82 (61)

A: abnormal: carcinoma only; B: abnormal: carcinoma or high-grade squamous intraepithelial lesion only; C: abnormal: carcinoma, high-grade squamous intraepithelial lesion, or low-grade squamous intraepithelial lesion; D: abnormal: carcinoma, high-grade squamous intraepithelial lesion, low-grade squamous intraepithelial lesion, or unsatisfactory; E: abnormal: carcinoma, high-grade squamous intraepithelial lesion, low-grade squamous intraepithelial lesion, unsatisfactory, or atypical squamous cells/atypical glandular cells.

the validation slides ( $P < 0.001$ , Kruskal–Wallis non-parametric test). The litigation cases were found to be significantly more likely to be interpreted as negative or as ASC/AGC compared with the validation cases ( $P < 0.001$  for each, Fisher exact test).

The ease of interpretation for the litigation slides as a whole (i.e., the percentage of the time that the group as a whole was termed abnormal [including “unsatisfactory”]) was 56% (all diagnoses except NIL from Table 1). Of these, there were 14 litigation cases (10%) that were consistently interpreted as abnormal (any interpretation other than NIL) by all 10 cytotechnologists. If these 14 cases were removed, the ease of interpretation for the group was reduced to 51%.

Cases that would be diagnosed regularly and reliably at different interpretation thresholds (i.e., ASC/AGC, low-grade squamous intraepithelial lesion [LSIL], high-grade intraepithelial lesion [HSIL], etc.) and different result thresholds (i.e., identified 7 of 10 times, 8 of 10 times, etc.) for the litigation slides and validation slides are shown in Tables 3 and 4. The inclusion of ASC/AGC cases as abnormal in the litigation cases significantly increases the number of cases that are deemed regularly and reliably detected at all performance thresholds ( $P \leq 0.032$  for all, Fisher exact test). That is, the number of cases that are determined to be regularly and reliably detected as abnormal is significantly greater at each threshold (e.g., the threshold of 8 of 10) when compared with the number of cases determined to be regularly and reliably detected when interpretations of ASC/AGC that were not counted as abnormal at that same threshold. In contrast, for the validation cases, the inclusion of ASC/AGC cases was found to significantly increase the number of cases deemed to be regularly and reliably

**TABLE 4**  
**Number and Percentage of Slides Regularly and Reliably Detected Using Different Thresholds for Validation Slides (n = 122)**

No. of interpretations	A	B	C	D	E
10 of 10	6 (5)	46 (38)	73 (60)	73 (60)	95 (78)
9 of 10	8 (7)	55 (45)	87 (71)	89 (73)	110 (90)
8 of 10	12 (10)	66 (54)	95 (78)	97 (80)	114 (93)
7 of 10	14 (11)	69 (57)	104 (85)	104 (85)	115 (94)
6 of 10	18 (15)	81 (66)	114 (93)	114 (93)	121 (99)
5 of 10	25 (20)	86 (70)	119 (98)	119 (98)	122 (100)

A: abnormal: carcinoma only; B: abnormal: carcinoma or high-grade squamous intraepithelial lesion only; C: abnormal: carcinoma, high-grade squamous intraepithelial lesion, or low-grade squamous intraepithelial lesion; D: abnormal: carcinoma, high-grade squamous intraepithelial lesion, low-grade squamous intraepithelial lesion, or unsatisfactory; E: abnormal: carcinoma, high-grade squamous intraepithelial lesion, low-grade squamous intraepithelial lesion, unsatisfactory, or atypical squamous cells/atypical glandular cells.

**TABLE 5**  
**The Chance of a Case Being Classified Based on Different Thresholds and Known Ease of Interpretation**

No. of interpretations	Ease of interpretation				
	40%	50%	80%	99%	99.9%
10 of 10	< 0.01	< 0.01	11	90	100
9 of 10	< 0.01	0.01	27	9	< 0.01
8 of 10	0.01	4	30	< 0.01	< 0.01

detected for performance thresholds of 8 of 10 and higher ( $P \leq 0.023$  for all, Fisher exact test). Of the litigation slides, 10% were interpreted as at least ASC/AGC by all observers, in contrast to 78% of the validated cases ( $P < 0.001$ , Fisher exact test). Conversely, 6 of the litigation slides (4%) and none of the validation slides were deemed to be within normal limits by all observers ( $P = 0.03$ , Fisher exact test). A total of 120 of the 135 litigation cases (89%) were interpreted as NIL by at least 1 observer, compared with 24 of the validation cases (20%) ( $P < 0.001$ , Fisher exact test).

Although 78% of the validation cases were interpreted as ASC/AGC or higher by all observers, it was noted that the majority of cases interpreted as NIL were seen in the beginning of the program. Of the final 50 validation cases used in the program, 100% were interpreted as ASC/AGC or higher by all observers. This is significantly greater than for the preceding 72 cases ( $P < 0.001$ , Fisher exact test).

Estimates of how cases with different known eases of interpretation<sup>12</sup> would be classified are shown in Table 5. In brief, this table shows what the expected performance of cases with known levels of difficulty would be. For example, a case with a known ease of

interpretation of 50% would be expected to be found to be abnormal in 8 of 10 reviews: 4% of the time that it was reviewed by the panel. In contrast, a case with a known ease of interpretation of 80% would be expected to have the same result 30% of the time. Using this information, one can estimate how often a case of a particular ease of interpretation might be correctly classified and misclassified by the review process, depending on the specific threshold used to make this distinction.

As a surrogate for precision, the distribution of interpretations within each interpretation was examined using analysis of variance for nonparametric data. Specifically, we compared how many times an interpretation of ASC/AGC was made (10 of 10 times, 9 of 10 times, 8 of 10 times, etc.) with how many times an interpretation of HSIL was made (10 of 10 times, 9 of 10 times, etc.). A Kruskal-Wallis test demonstrated that for the litigation slides, the interpretation of ASC/AGC was significantly different from that of unsatisfactory, LSIL, and carcinoma (all  $P < 0.05$ ), and no statistical difference was noted between the interpretation of ASC/AGC and the interpretation of HSIL ( $P = 0.12$ ). In contrast, for the validation slides, the interpretation of ASC/AGC and unsatisfactory were both found to be significantly different from LSIL, HSIL, and carcinoma (all  $P < 0.001$ ).

To determine whether cytotechnologists who diagnosed validation cases as negative were more likely to diagnose litigation cases as negative, we compared the percentage of litigation cases that were diagnosed as NIL by cytotechnologists who diagnosed any validation case as NIL with the number of litigation cases that were interpreted as NIL by cytotechnologists who did not interpret any validation cases as NIL. A two-tailed Fisher exact test demonstrated no statistically significant difference ( $P = 0.12$ ).

## DISCUSSION

The data presented in the current study can be interpreted on several different levels.

First, litigation cases are significantly more likely to be diagnosed as negative or ASC/AGC than as unsatisfactory, LSIL, HSIL, or carcinoma. In addition, at every interpretation and performance threshold, litigation slides are significantly less likely to be regularly and reliably identified as abnormal than validation slides. Put simply, the litigation cases are significantly more difficult to identify and are less easy to interpret as abnormal than the biopsy-proven validation Pap slides. The ease of interpretation of the entire litigation group is 56%. If the cases that are always identified as ASC/AGC or higher are excluded, the percentage drops to 51%. This is slightly higher than the

estimated ease of interpretation of 40% for routine false-negative cases as a group.<sup>12</sup> This suggests that these cases did not fall beneath the standard for care. It is interesting to note that because the ease of interpretation is higher than expected, this suggests that at least some of these litigation cases are easier to diagnose than routine false-negative cases.

Second, the results appear to be independent of the individual cytotechnologist screening the case. It is reasonable to question whether the results could be affected by using cytotechnologists who have a very high threshold for abnormality. If this were so, one would expect that cytotechnologists who called the validation cases negative also would be more likely to interpret the litigation cases as negative. However, this was not the case. We were unable to show that litigation cases were more likely to be diagnosed as negative by cytotechnologists who interpreted validation cases as negative compared with those who did not.

Third, the data suggest that the interpretation of ASC/AGC may be as valid as the interpretation of HSIL for litigation cases. Using the distribution of interpretations as a surrogate for precision, for litigation cases the interpretation of ASC/AGC was distributed in a manner similar to HSIL and different from LSIL and carcinoma. In contrast, for validation cases, the interpretation of ASC/AGC was distributed differently from LSIL, carcinoma, and HSIL. The simplest interpretation of this result is that HSIL is comprised of two distinct categories: those that are easy to interpret and those that are difficult. The validation cases are comprised of HSIL cases that are easy to interpret and are distributed differently from cases of ASC/AGC. In the litigation group, the HSIL cases are just as difficult to interpret and are just as nonreproducible as the ASC/AGC cases. Because of this, an interpretation of ASC/AGC may not need to be excluded from the interpretation unless one is also willing to exclude an interpretation of HSIL. This result essentially recommends that cases of ASC/AGC be included as abnormal interpretations when using this method.

Why have cytologists been unwilling to accept an interpretation of ASC/AGC as abnormal in the context of blinded review? Although no one doubts that ASC/AGC is a valid and abnormal interpretation in the clinical context, some interpreters of blinded review have been unwilling to include responses of ASC/AGC as truly abnormal in interpreting the results of blinded review. The most common reason given is that the interpretation is not reproducible. The practice is similar to that used in some cytologic-histologic correlation settings in which a difference of interpretation of at least two steps is required to qualify as an error. This practice is used to ensure that the error one finds is

truly an error (i.e., is reproducible). Essentially, this is just a surrogate for reproducibility. One can use the data presented in Tables 3 and 4 to examine exactly how well requiring a two-step difference can mimic reproducibility. However, in the current setting there is no need to use a surrogate for reproducibility. Blinded review is already directly examining each case multiple times, so one knows exactly how reproducible each interpretation is. Therefore, although requiring a two-step difference makes sense as a surrogate for reproducibility in a setting in which one is only examining one case at a time, it does not make sense in the current study, in which reproducibility can be measured easily.

Fourth, the data do help us decide which performance threshold should be used as regularly and reliably identified. There are some cases that everyone would agree are regularly and reliably detected and others that are not. Of the litigation cases, 10% were diagnosed as at least ASC/AGC by all observers and 8% were diagnosed as NIL by all observers. Of the validation slides, 78% were diagnosed as ASC/AGC by all observers and none was diagnosed as NIL by all observers. Unfortunately, many cases fall in between these two results, and not every observer may be in agreement regarding the best categorization for these cases. Defining a justifiable, logical, and scientifically based rationale for making the decision in these cases is critical.

Fifth, as we have shown previously, there are two obvious performance thresholds that might be used to deem a case as regularly and reliably identified: 7 of 10 and 10 of 10.<sup>12</sup> Multiple-slide, blinded review is similar to any other test, and standards must be set for the interpretation of the results. Ideally, these standards will be straightforward and justifiable. The 10 of 10 threshold is relatively easy to understand. Specifically, the rationale for this is that practicing cytotechnologists should not be required to always detect cases that cannot always be detected. If, for example, a case can only be recognized as abnormal 80–90% of the time, then it is unreasonable to require that a practicing cytotechnologist identify this case 100% of the time. If other cytotechnologists also fail to render the slide as abnormal, then the case cannot be regularly and reliably detected. The results of the current study demonstrated that 10% of the litigation cases examined were of this type.

However, only 78% of the validation cases met this criteria. In fact, this is nearly exactly the same error rate as was reported for Pap slides in general.<sup>13,14</sup> This strongly suggests that, as a whole, the “validation” cases used in the current study are not cases that can always be identified, but simply regular ordinary rou-

tine cases. The ease of interpretation of these groups is not 100%, but 78%. It is interesting to note that, as the program went on, this appeared to be recognized, and in the last 50 cases the ease of interpretation had increased to 100%. It would appear that the program may be improved by more formally defining what ease of interpretation one requires for these validation cases, and explicitly defining the criteria needed for cases to achieve this, as is reflected in other programs such as the College of American Pathologists Gynecologic Cytology Program.<sup>15</sup>

The 7 of 10 threshold is based on results in the literature defining expected false-negative rates. Specifically, this threshold is chosen to ensure that the chance of a case being misclassified is the same for both groups (i.e., for cases that should be regularly and reliably identified and for cases that should *not* be regularly and reliably identified). To determine the chance that a case will be misclassified, one must know the population of cases that is being reviewed. In our previous work,<sup>12</sup> we assumed that the cases that were being reviewed were the regular and routine cases that are reviewed in the cytology laboratory. The threshold of 7 of 10 is designed to ensure that for the typical population of cases, the chance of misclassifying a case as regularly and reliably identified is the same as the chance of misclassifying a case as *not* regularly and reliably identified. This typical population of cases is defined based on consensus review for all abnormal cases.

However, we have already shown that the litigation cases are much more difficult to classify as abnormal than the validation cases, which are similar to routine cases in the literature. Therefore, the 7 of 10 threshold is no longer valid because it applies to a review of a different population of cases. Because the group of cases is more difficult to classify than the routine cases, one can expect that there are relatively few cases that can be identified regularly and reliably in this group, and therefore, at this threshold, the chance of misclassifying a case as regularly and reliably detected is much higher than the chance of misclassifying a case as *not* regularly and reliably detected.

This chance of a case being misclassified into each of these categories can be further estimated using data published previously.<sup>12</sup> These data are summarized in Table 5. With only 10 reviews, the test is relatively underpowered. As a result, although it performs very well at distinguishing between cases that are always identified (ease of interpretation of 99.9%) and typical false-negative cases (ease of interpretation of 40%), it is not able to reliably distinguish cases in between (routine cases, with an ease of interpretation of 80%)

and cases that are almost always identified (ease of interpretation of 99%). Using thresholds of 8 of 10 and 9 of 10 are much more likely to misclassify a case as regularly and reliably identified than the opposite. Using 10 of 10 as a threshold, 11% of routine cases (ease of interpretation of 80%) will simply by chance be identified as regularly and reliably identified. Cases that are almost always identified (ease of interpretation of 99%) will be misclassified as *not* regularly and reliably identified 10% of the time using a threshold of 10 of 10. Therefore, using a 10 of 10 threshold in the current study results in miscategorizing approximately 10% of moderate to difficult cases (ease of interpretation of 80% and 99%, respectively) and only 1% of the litigation cases. This suggests that of the 14 cases that were identified all 10 times, no greater than 10% (i.e., 1 to 2 cases) were misclassified.

Where does this analysis leave an "expert" cytologist who still believes that they can determine whether a case is regularly and reliably detectable based on a single review? The data in this report tells us exactly. Referring to Table 1, based on all the reviews of all cases, an average of 56% of the litigation cases were determined to be something other than NIL. This means that for this population of cases, on average, a single "expert" cytologist will determine that 56% of these cases are abnormal, even though only 10% clearly are. Although specific individual reviewers will obviously have variable results, this finding in fact is very similar to that of the consensus panels in use in double-blind crossover studies.<sup>13,14</sup> In these studies, the vast majority of cases identified as abnormal by only one reviewer were not deemed abnormal after consensus review. Of course, there may be other ways of evaluating these cases. Specific cytologic criteria with which to distinguish cases that can be identified regularly and reliably by practicing cytologists have been published previously.<sup>16-21</sup> Although only recently defined, to our knowledge not a single expert to date has referred to these criteria when determining whether a slide is considered abnormal or has been originally diagnosed beneath the standard of care when they make their decision.

At this point, the data in the literature are abundant, and the wealth of information regarding false-negative rates in practice and in blinded review has been analyzed extensively. We can predict and compare the performance of laboratories, of different interpretations, of individuals, and of individual cases. The results in the current study are consistent with those predicted based on analysis of the system.<sup>12</sup> The science of false-negative rates and blinded review has been extensively and almost fully characterized. For those interested in predicting performance or out-

come (the essence of science) in the clinical setting or in multiple-slide, blinded review, the literature provides more than sufficient information with which to make highly accurate predictions and analyze the results. In our opinion, there is no longer any excuse for "experts" to interpret the expected outcome of individual slides without basing that interpretation on this wealth of scientific information.<sup>22</sup>

This program clearly identified at least 3 different groups of cases with different eases of interpretation: litigation cases (ease of interpretation of 56%), routine validation cases (ease of interpretation of 78%), and true validation cases (ease of interpretation of 100%). The ease of interpretation of the litigation cases was higher than that in the literature for typical false-negative cases, suggesting that there are cases in this group that are easier to detect than one would expect. Approximately 10% of the litigation cases brought to blinded review in this program are detected regularly and reliably. Approximately 4% of cases are universally interpreted as negative. The program may be improved by considering ASC/AGC as an abnormal interpretation for the purposes of evaluating the results of multiple-slide, blinded review and explicitly defining criteria for the selection of validation cases. A threshold of 10 of 10 appears to be the most logical threshold to use and may result in a misclassification rate of 10%. The results suggest that experts who rely on a single review will inevitably overestimate by a large margin (more than four times greater) the percentage of litigation cases that they state should have been regularly and reliably identified originally by the cytotechnologist or pathologist when the slide was initially interpreted in the clinical setting.

## REFERENCES

1. Fitzgibbons PL, Austin RM. Expert review of histologic slides and Papanicolaou tests in the context of litigation or potential litigation. Surgical Pathology Committee and Cytopathology Committee of the College of American Pathologists. *Arch Pathol Lab Med.* 2000;124:1717-1719.
2. Volk EE, Austin RM. Before litigation, blinded rescreening a must. CAP updates guide for review of Pap tests in litigation or potential litigation. *CAP Today.* 2002.
3. Austin RM. Updated ASC/AGC pap litigation guidelines support unbiased blinded rescreening as key pre-litigation test. *ASC/AGC Bulletin.* 2002;63-68.
4. Austin RM. Results of blinded rescreening of Papanicolaou smears versus biased retrospective review. *Arch Pathol Lab Med.* 1997;121:227-228.
5. The Pap smear on trial. *Lab Med.* 1997;26.
6. Frable WJ. Guidelines for experts reviewing Papanicolaou smear litigation cases. *Arch Pathol Lab Med.* 1997;121:331-334.
7. Stastny JF, Spires S, Austin RM. Pap litigation: where do we go from here? *Diagn Cytopathol.* 1998;19:235-237.

8. Austin RM. The Papanicolaou smear. Medicine's most successful cancer screening procedure is threatened. *JAMA*. 1997;277:754–755.
9. Frable WJ, Austin RM, Greening SE, et al. Medicolegal affairs. International Academy of Cytology Task Force summary. Diagnostic Cytology Towards the 21st Century: an International Expert Conference and Tutorial [review]. *Acta Cytol*. 1998;42:76–119; discussion 120–132.
10. Holladay EB, Austin RM. Multiple slide blinded rescreening for review of litigation slides. In: Allen K, Holladay EB, editors. Risk management for the cytology laboratory. 2001: 149–162.
11. Solomon D, Nayar R. The Bethesda system for reporting cervical cytology. 2nd edition. New York: Springer-Verlag, 2004.
12. Renshaw AA, Young ML, Holladay EB. Blinded review of Pap smears in the context of litigation: using statistical analysis to define appropriate thresholds. *Cancer (Cancer Cytopathol)*. 2004;102:136–141.
13. Renshaw AA, Lezon KM, Wilbur DC. The human false-negative rate of rescreening Pap tests. Measured in a two-arm prospective clinical trial. *Cancer (Cancer Cytopathol)*. 2001; 93:106–110.
14. Renshaw AA. Measuring sensitivity in gynecologic cytology. A review. *Cancer (Cancer Cytopathol)*. 2002;96:210–217.
15. College of American Pathologists. Interlaboratory Comparison Program in Gynecologic Cytopathology (PAP) Year End Summary Report. Chicago: College of American Pathologists, 2000.
16. Renshaw AA, Dubray-Benstein B, Cobb CJ, et al. Cytologic features of squamous cell carcinoma in ThinPrep slides. *Arch Pathol Lab Med*. 2004;128:403–405.
17. Renshaw AA, Mody DR, Lozano RL, Volk EE, Davey DD, Birdsong GG. Adenocarcinoma in situ of the cervix is significantly more difficult to identify and categorize than adenocarcinoma, high grade squamous intraepithelial lesion, or squamous cell carcinoma. *Arch Pathol Lab Med* 2004;128: 153–157.
18. Renshaw AA, Schulte MA, Plott E, et al. Cytologic features of high-grade squamous intraepithelial lesion in ThinPrep Papanicolaou test slides: comparison of cases that performed poorly with those that performed well in the College of American Pathologists Interlaboratory Comparison Program in Cervicovaginal Cytology. *Arch Pathol Lab Med*. 2004;128:746–748.
19. Renshaw AA, Prey MU, Moriarty AT. Cytologic features of high grade squamous intraepithelial lesion in conventional slides: comparison of cases which performed poorly and well in the College of American Pathologists gynecologic cytology program. *Arch Pathol Lab Med*. 2005;129:733–735.
20. Renshaw AA, Dubray-Benstein BL, Hughes JL. Cytologic features of low grade squamous intraepithelial lesion in ThinPrep and conventional smear specimens. *Arch Pathol Lab Med*. 2005;129:23–25.
21. Renshaw AA, Hughes JL. Cytologic features of squamous cell carcinoma in conventional smears. *Arch Pathol Lab Med*. In press.
22. Renshaw AA. Experts in Wonderland: In search of the right test and the scientific method. *Diagn Cytopathol*. 2000; 23:297–298.