

Review of the Bethesda System Atlas Does Not Improve Reproducibility or Accuracy in the Classification of Atypical Squamous Cells of Undetermined Significance Smears

Ann E. Smith, M.D.¹
 Mark E. Sherman, M.D.¹
 David R. Scott, M.D.²
 Sana O. Tabbara, M.D.³
 Lawrence Dworkin, M.D.²
 June Olson, M.D.⁴
 John Thompson, M.D.²
 Christine Faser, M.D.⁵
 Joy Snell, M.D.⁶
 Mark Schiffman, M.D., M.P.H.⁷

¹ Department of Pathology, Johns Hopkins Medical Institutions, Baltimore, Maryland.

² Regional Laboratory, Kaiser Foundation Health Plan-Northwest, Clackamas, Oregon 97015.

³ Department of Pathology, The George Washington University Medical Center, Washington, DC.

⁴ Department of Pathology, Portland Providence Medical Center, Portland, Oregon.

⁵ Pathology Laboratories, Muskegon, Michigan.

⁶ Department of Pathology, Comanche County Memorial Hospital, Lawton, Oklahoma.

⁷ National Cancer Institute, Bethesda, Maryland.

Presented in part at the 87th annual meeting of the United States and Canadian Academy of Pathology, Boston, Massachusetts, March 1–6, 1998.

Address for reprints: Ann E. Smith, M.D., Department of Pathology, Johns Hopkins Medical Institutions, Carnegie 400, 600 N. Wolfe Street, Baltimore, MD 21287.

Received September 8, 1999; revision received February 11, 2000; accepted February 22, 2000.

BACKGROUND. The Bethesda System (TBS) and its accompanying atlas were developed to promote uniform diagnosis and reporting of cervical and vaginal cytology, especially with respect to borderline abnormal smears. The authors assessed whether group study of TBS atlas improves the reproducibility and accuracy of the cytopathologic diagnosis of equivocal Papanicolaou smears.

METHODS. One hundred “atypical” smears were divided into pretest and posttest sets containing equal numbers of negative, atypical squamous cells of undetermined significance (ASCUS), and squamous intraepithelial lesion (SIL) diagnoses based on a five-member panel review. Two comparable teams of four pathologists from George Washington University Medical Center (Washington, DC) and Kaiser Permanente (Portland, OR), each comprised of two more experienced cytopathologists and two less experienced pathologists, independently reviewed the 50 pretest slides and classified the slides according to TBS as negative, ASCUS, or SIL. The teams then conducted group study sessions using TBS atlas. After the review, the pathologists independently classified the 50 posttest slides in a similar manner.

RESULTS. Pretest, pair-wise interobserver agreement ranged from 30% to 66% compared with 34–62% for posttest agreement. Absolute percent agreement of reviewers’ diagnoses with a previously developed consensus diagnosis based on opinions of a five-expert panel (cytopathologic certainty scale) ranged from 44% to 62% for the pretest set and from 40% to 60% for the posttest set. Comparison of the detection of oncogenic human papilloma virus (HPV) DNA by hybrid capture tube test with smears classified as negative, ASCUS, or SIL revealed that seven of eight reviewers did not demonstrate a stronger association between HPV detection and cytologic diagnosis in the posttest set.

CONCLUSIONS. Review of TBS atlas by itself does not appear to improve the reproducibility or accuracy of cytologic diagnoses. The lack of improvement was similar among the pathologists involved regardless of experience level or whether they had a close working relation. *Cancer (Cancer Cytopathol)* 2000;90:201–6.

© 2000 American Cancer Society.

KEYWORDS: atypical squamous cells of undetermined significance, Bethesda System, Papanicolaou, cervicovaginal, cytopathology, interobserver agreement, reproducibility, accuracy.

In December 1988, a National Cancer Institute-sponsored workshop developed the Bethesda System (TBS) for reporting cervical and vaginal cytologic diagnoses.¹ TBS was developed to provide uniform terminology for reporting cytologic diagnoses, but specific morphologic criteria were not proposed for the diagnostic categories created.²

In April 1991, a second workshop convened to resolve difficulties encountered during the early implementation of TBS and to develop specific criteria for TBS diagnostic categories.³ Despite improvements in TBS, the category of "atypical squamous cells of undetermined significance" (ASCUS) remained problematic for practicing pathologists. Notably, studies demonstrated that interobserver agreement was poor, that pathologists diagnosed ASCUS with varying frequency, and that the likelihood of an underlying squamous intraepithelial lesion (SIL) also differed among reviewers.

Accordingly, three measures were taken to increase the clinical utility of TBS. First, the National Cancer Institute (NCI) developed interim guidelines for the management of abnormal cervical cytology.⁴ Second, the NCI launched a large, multidisciplinary trial, The ASCUS LSIL Triage Study (ALTS), to determine the best management of minimally abnormal Papanicolaou (Pap) smears.⁵ Third, the Bethesda System atlas⁶ was created to illustrate pathologic entities meeting the criteria set forth in the revised TBS. The present study was conducted to evaluate whether the TBS atlas is effective in achieving one of its major functions, to improve the reproducibility and accuracy of the cytopathologic diagnosis of equivocal Pap smears.

MATERIALS AND METHODS

Case Selection

This study included smears that were obtained from 100 participants in a 5-year prospective cohort study of human papilloma virus (HPV) infection and cervical SIL sponsored by the NCI that was conducted at Kaiser Permanente (Portland, OR).^{7,8} The 100 smears represent a stratified, random sample of a larger set of 200 smears that have been characterized previously in detail.⁹

Briefly, the original study set consisted of 200 smears that originally were classified as "atypical" using old nomenclature. These smears were independently reevaluated by a masked, five-member panel using TBS and correlated with HPV DNA detection using multiple methods. The consensus diagnosis of the pathology panel for each slide was expressed as a "cytologic certainty score"⁸ indicating the likelihood of SIL. The cytologic certainty score was ascertained by assigning a value of 0.0 to each pathologist's diagnosis of negative, 0.5 to diagnoses of ASCUS, and 1.0 to diagnoses of SIL. Therefore, the score for the slides ranged from 0.0 (unanimous diagnosis of normal by the five-member panel) to 5.0 (unanimous diagnosis of SIL). Overall, a score of 0–1.0 was classified as normal, a score of 1.5–3.0 was classified as equivocal,

and a score of 3.5–5.0 was classified as likely SIL. The 100 slides that were used in the current report were identified by taking a stratified, random sample across the range of cytologic certainty scores assigned by the original pathology panel to each of the 200 slides. The 100 slides were then divided into comparable pretest and posttest sets that were comprised of 50 slides each with similar cytologic certainty scores.

In the original review of the 200 "atypical" smears, cytopathologic diagnoses were strongly related to the detection of oncogenic HPV types, but the detection of low risk types was not strongly related.⁹ Accordingly, only the detection of oncogenic HPV types is presented in this report. HPV detection using a polymerase chain reaction-based method, a Southern hybridization technique, and the U.S. Food and Drug Administration-approved, first generation of the hybrid capture system HPV DNA test (HCT) produced similar conclusions.¹⁰ Therefore, only the hybrid capture results for oncogenic types are presented for simplicity.

Cytology Review

Two comparable teams of four pathologists each were selected from George Washington University Medical Center (Washington, DC) and Kaiser Permanente (Portland, OR). Each team was comprised of two experienced cytopathologists and two less experienced pathologists. The "more experienced" participants were board certified in cytopathology and were routinely involved in clinical cytopathology at their respective institutions. The "less experienced" reviewers were either completing residency training in anatomic pathology (George Washington University Medical Center) or practicing pathologists who did not concentrate on cytopathology in their pathology practice (Kaiser Permanente). First, each pathologist, who was masked to the existing data base, independently reviewed a pretest set of 50 smears and classified the slides according to TBS as negative, ASCUS, or SIL. Then, the teams conducted group study sessions using TBS atlas, focusing on the cytopathologic distinctions between negative, ASCUS, and SIL. The review sessions were based on group review by the test subjects themselves, without an independent outside expert. The two teams conducted their respective reviews during either two sessions over a period of 3 weeks (George Washington University Medical Center) or in a single session (Kaiser Permanente), at which time the team members discussed each published illustration of TBS atlas, page by page, without microscopic examination of slides. After the review sessions, the pathologists independently reviewed the posttest set of 50 smears and classified these using TBS.

TABLE 1
Comparison of Interobserver Agreement Before (Pretest) and After (Posttest) Review of The Bethesda System Atlas

Reviewers	Pretest		Posttest	
	Agreement (%)	Kappa (weighted) ^a	Agreement (%)	Kappa (weighted) ^a
More experienced	54–66	0.43–0.60	54–62	0.35–0.48
Same institution	62–66	0.55–0.60	54	0.35–0.43
Different institution	54–64	0.43–0.55	54–62	0.41–0.48
More vs. less experienced	30–60	–0.07–0.51	34–54	–0.03–0.40
Same institution	30–60	–0.07–0.51	36–52	0.09–0.40
Different institution	36–58	0.10–0.44	34–54	–0.30–0.40
Less experienced	34–54	0.02–0.45	36–54	0.06–0.32
Same institution	34–44	0.02–0.23	40–50	0.11–0.27
Different institution	36–54	0.04–0.45	36–54	0.06–0.32

^a Kappa (weighted) ≥ 0.75 , “excellent” agreement beyond chance; <0.75 and >0.40 , “fair-to-good” agreement beyond chance; ≤ 0.40 , “poor” agreement beyond chance.

Statistical Analysis

The utility of TBS atlas in improving cytologic diagnosis was assessed by comparing the pretest performance with the posttest performance of the panel with respect to interobserver agreement and to diagnostic accuracy using the cytologic certainty score of the original panel⁹ and HPV testing¹⁰ as “gold standards.” Interobserver agreement or reproducibility was analyzed by comparing all pair-wise combinations of reviewers, both within and between experience levels and institutions. Standard three-by-three contingency tables were constructed for each pair of pathologists to calculate and compare pretest and posttest overall percent agreement for TBS categories using standard independent Z-tests. To correct for the level of expected agreement due to chance, a weighted “kappa” statistic was calculated for each pair of reviewers that adjusted the observed agreement for the expected level of agreement predicted to occur just by chance. The kappa statistic has a theoretic maximal value of 1.0, which would indicate complete agreement. Commonly, a value for kappa ≥ 0.75 is interpreted as “excellent” agreement beyond chance, kappa ≤ 0.40 signifies “poor” agreement, and kappa values in the range of 0.40–0.75 indicates “fair-to-good” agreement (with 0.0 representing no association and –1.0 indicating complete disagreement). The kappa analysis was “weighted” to quantify the relative magnitude of disagreement. Exact agreement was given a maximal weight of 1.0, and all disagreements were given less than maximal weight on a gradation of 0.5 per step of disagreement, such that SIL versus normal was assigned a weight of 0.0. The interpretation of the magnitude of the weighted kappa is the same as that of the unweighted kappa.¹¹ If reproduc-

ibility improved after training with TBS atlas, then the kappa value should have increased.

The change in diagnostic accuracy for each pathologist after TBS study was evaluated first by comparing the pretest and posttest diagnoses with the five-point “cytopathologic certainty scale.”⁹ The overall percent agreement was calculated for diagnoses of negative, equivocal, or SIL. For each individual pathologist, the overall percent disagreement was defined as a diagnosis of negative or SIL that was assigned as the opposite extreme (SIL or negative, respectively) by the expert panel. If accuracy was improved after training with TBS atlas, then the overall percent agreement between our pathologists’ diagnoses and the cytologic certainty scale should have increased predictably, whereas the percent disagreement should have decreased.

The change in pretest and posttest accuracy of each pathologist also was assessed by comparing the detection of oncogenic HPV DNA by HCT¹⁰ in women with smears classified as negative, ASCUS, or SIL. If review of TBS atlas improved diagnostic accuracy, then one would expect a stronger association between HPV detection and cytologic diagnoses in the posttest set compared with the pretest set using the Mantel–Haenszel chi-square test for trend.¹² Specifically, the increased detection of HPV associated with SIL compared with ASCUS and in detection of ASCUS compared with normal would improve after training with the atlas.

RESULTS

The pretest, pair-wise, interobserver percent agreement ranged from 30% to 66% compared with 34–62% for posttest agreement (Table 1). Overall, regardless of experience level or institution, there was no significant improvement in interobserver agreement after review

TABLE 2
Comparison of Absolute Percent Agreement and Disagreement (Normal vs. Squamous Intraepithelial Lesion) of Reviewers' Pretest and Posttest Diagnoses with the "Cytopathologic Certainty Scale"

Reviewer ^a	Pretest (%)	Posttest (%)
A		
Agreement	52	52
Disagreement	6	2
B		
Agreement	56	52
Disagreement	4	0
C		
Agreement	54	60
Disagreement	8	4
D		
Agreement	46	54
Disagreement	12	16
E		
Agreement	62	52
Disagreement	0	10
F		
Agreement	50	46
Disagreement	0	10
G		
Agreement	44	40
Disagreement	18	10
H		
Agreement	58	46
Disagreement	2	6

^a Reviewers A, B, E, and F were more experienced, and reviewers C, D, G, and H were less experienced.

with TBS atlas. The corresponding weighted kappa values ranged from -0.07 to 0.60 for the pretest set compared with a range from -0.03 to 0.48 for the posttest set. Only the more experienced cytopathologists had consistent "fair-to-good" agreement by kappa analysis. All of the less experienced pairs and the more experienced versus less experienced pairs had "poor" agreement after review with TBS atlas.

The percent absolute agreement between the reviewers' diagnoses and the diagnoses of the cytopathologic certainty scale ranged from 44% to 62% before review with TBS atlas and from 40% to 60% after atlas review (Table 2). The percent of disagreement (negative vs. SIL) ranged from 0% to 18% for the pretest set of smears and from 0% to 16% for the posttest set of smears. Regardless of experience level, there was no consistent or significant change in the percent agreement or disagreement after training with TBS atlas.

A comparison of the detection of oncogenic HPV DNA by HCT¹⁰ in smears classified as negative, ASCUS, or SIL by the reviewers both before and after review with the atlas is illustrated in Table 3. For nearly all of the reviewers, there was a higher rate of

TABLE 3
Comparison of Pretest and Posttest Diagnoses with the Detection of Oncogenic Human Papilloma Virus DNA by Hybrid Capture Tube Test

Reviewer ^a / Diagnosis	Pretest (n = 50)	Positive HCT (%) ^b	Posttest (n = 50)	Positive HCT (%) ^b
A				
Negative	23	4 (17)	21	2 (10)
ASCUS	17	8 (47)	13	7 (54)
SIL	10	7 (70)	16	12 (75)
Chi-square ^c	8.853 (<i>P</i> = 0.003)	—	16.137 (<i>P</i> < 0.01)	—
B				
Negative	19	3 (16)	19	1 (5)
ASCUS	18	6 (33)	20	13 (65)
SIL	13	10 (77)	11	7 (64)
Chi-square ^c	11.452 (<i>P</i> = 0.001)	—	12.272 (<i>P</i> < 0.01)	—
C				
Negative	19	4 (21)	18	4 (22)
ASCUS	11	3 (27)	10	4 (40)
SIL	20	12 (60)	22	13 (59)
Chi-square ^c	6.196 (<i>P</i> = 0.013)	—	5.432 (<i>P</i> = 0.020)	—
D				
Negative	24	7 (29)	17	7 (41)
ASCUS	15	6 (40)	21	8 (38)
SIL	11	6 (55)	12	6 (50)
Chi-square ^c	2.042 (<i>P</i> = 0.153)	—	0.171 (<i>P</i> = 0.679)	—
E				
Negative	18	1 (6)	25	6 (24)
ASCUS	15	5 (33)	17	8 (47)
SIL	17	13 (76)	8	7 (88)
Chi-square ^c	18.225 (<i>P</i> < 0.01)	—	9.793 (<i>P</i> = 0.002)	—
F				
Negative	18	0 (0)	25	5 (20)
ASCUS	19	10 (53)	12	6 (50)
SIL	13	9 (69)	13	10 (77)
Chi-square ^c	16.203 (<i>P</i> < 0.01)	—	11.547 (<i>P</i> = 0.001)	—
G				
Negative	19	7 (37)	12	4 (33)
ASCUS	17	7 (41)	19	8 (42)
SIL	14	5 (36)	19	9 (47)
Chi-square ^c	0.001 (<i>P</i> = 0.971)	—	0.569 (<i>P</i> = 0.451)	—
H				
Negative	12	1 (8)	14	3 (21)
ASCUS	20	6 (30)	18	8 (44)
SIL	18	12 (67)	18	10 (56)
Chi-square ^c	10.802 (<i>P</i> = 0.001)	—	3.594 (<i>P</i> = 0.058)	—

HCT: hybrid capture tube test; ASCUS: atypical squamous cells of undetermined significance; SIL: squamous intraepithelial lesion.

^a Reviewers A, B, E, and F were more experienced, and reviewers C, D, G, and H were less experienced.

^b Oncogenic human papilloma virus DNA by HCT test.

^c Mantel-Haenzel chi-square test for trend (1 degree of freedom).

HPV positivity in smears that were classified as SIL compared with smears that were classified as negative. However, using the Mantel-Haenzel chi-square test for trend, only one of the eight reviewers demonstrated a stronger association between HPV detection and cytologic diagnoses in the posttest set.

DISCUSSION

In this study, reproducibility measured as pair-wise, absolute percent interobserver agreement or as weighted kappa values, did not improve after study of TBS atlas. In addition, the accuracy of the pathologists' diagnoses based on a comparison with the consensus diagnosis of an expert panel⁹ or HPV testing¹⁰ showed no significant change after study of the atlas. Therefore, there is no evidence from this study that review of TBS atlas by itself improves reproducibility or accuracy of cytologic diagnoses. The lack of improvement in reproducibility and accuracy after TBS review was similar among our pathologists irrespective of their level of experience or whether they had a close working relation.

The poor reproducibility at baseline that we observed is similar to that previously reported. In a previous study using the same smears, five expert cytopathologists achieved complete consensus in less than 30% of the cases.⁹ Notably, none of the smears was classified unanimously as ASCUS by all five reviewers, and the smears that were reclassified as negative accounted for most cases of unanimous agreement. In 1992, the College of American Pathologists (CAP) Cytopathology Committee selected 13 Pap smears that originally were diagnosed as ASCUS in their laboratory. On rereview by each committee member, 8% of the rediagnoses were classified as normal, 20% were classified as reactive, 62% were classified as ASCUS, and 10% were classified as LSIL. In 1994, the CAP Cytopathology Committee reviewed 31 Pap smears that originally were diagnosed as ASCUS. These cases were presented as 35-mm slides to 17 members and experts. Only 7 of 31 smears were classified as ASCUS from greater than 70% of the members, and only 5 of 31 smears received a diagnosis of ASCUS from greater than 80% of the members.¹³ In 1997, Renshaw et al.¹⁴ chose 80 smears, 74 of which were classified originally as ASCUS. Each smear was reviewed by three observers and diagnosed as negative, reactive, ASCUS, LSIL, or HSIL. All ASCUS diagnoses were qualified further as favor reactive, ASCUS not otherwise specified, and ASCUS favor SIL. Complete agreement was seen in only 11% of smears and in only 2 cases that were diagnosed originally as ASCUS. In an attempt to increase interobserver agreement, the authors reduced the diagnostic categories to negative, ASCUS, and SIL, after which, the highest level of agreement had a kappa value of 0.37 (poor agreement).

With the poor level of interobserver agreement in the diagnosis of ASCUS, accuracy of the diagnosis becomes an important factor in terms of guiding management decisions. Based on our understanding of the

biology of oncogenic HPV types and cervical neoplasia,⁷ this study demonstrates that an ASCUS diagnosis can represent a wide spectrum of clinical entities from benign changes to cancer. Previous reports have shown similar variety in clinical outcomes of ASCUS based on follow-up studies. In 1994, Davey et al.¹⁵ reported follow-up data from ASCUS patients seen in four laboratories. Of the patients with ASCUS, follow-up data were available on 60–74% of patients. Of those with follow-up data, 50–69% had benign findings, 1–21% showed persistent ASCUS, and 10–43% had SIL with up to 6% having HSIL. In 1996, Howell and Davis¹⁶ reported follow-up data on 124 of 193 ASCUS smears. Of these, 124 follow-up smears revealed that 58% had no epithelial lesion, 12.9% had persistent ASCUS, and 29.1% had SIL. On rereview of the “ASCUS” smears, which showed no epithelial lesion on follow-up, only 19.5% were confirmed as consistent with ASCUS. Of the remainder, 75.1% showed inflammatory changes and/or metaplasia, 2.7% showed cells with enlarged nuclei, and 2.7% showed decreased estrogen effect. In 1997, Williams et al.¹⁷ published results of a correlative histologic and follow-up study on 668 smears that originally were diagnosed as ASCUS. Of those 668, 41% had biopsy follow-up. The biopsies showed condyloma in 36%, LSIL in 14%, and HSIL in 8%. None of the biopsies showed invasive carcinoma. Overall, the findings of our study support the fact that ASCUS is a highly heterogeneous diagnosis with a resulting low level of specificity. It is interesting to note that, although the more experienced reviewers appeared to achieve greater diagnostic accuracy compared with the results of hybrid capture testing (Table 3), this did not appear to be the case when reviewers' diagnoses were compared with those of the “cytopathologic certainty scale” (Table 2). These observations further emphasize the “uncertainty” of an ASCUS diagnosis. Even after intensive review of TBS atlas, none of our reviewers, most notably the less experienced cytopathologists, demonstrated significant improvement in the accuracy of their diagnoses on borderline smears.

Although TBS atlas provides precise criteria, the ASCUS category remains ill defined. Based on the results of this study, review of TBS atlas by itself does not increase interobserver agreement or accuracy of ASCUS diagnoses. However, in reality, TBS atlas is to be used as a training tool in conjunction with microscopy and other ancillary technology.¹⁸ The atlas alone cannot address the entire morphologic continuum, including artifact, variations in adequacy, and partial cellular obscuring, that typically is seen in Pap smear cytology. Furthermore, the diagnoses of equivocal smears, in reality, are made with some knowledge of

the patient's age, demographics, and previous Pap smear or clinical history. This study, therefore, is limited in its assessment of TBS atlas as an effective tool for improving the reproducibility and accuracy of interpreting borderline smears. In addition, a possible limitation of this study was our inability to control for exposure to TBS atlas at baseline. This concern would have been greater had the analysis shown excellent accuracy and reproducibility, but the relatively unimpressive results suggest at a minimum that review of the atlas does not promote excellent results. Furthermore, exposure to the atlas at the beginning of the study was likely to be least intense for the less experienced reviewers.

Ultimately, our understanding of the biology of HPV and cervical neoplasia will allow for subcategorization of ASCUS for the purpose of management decisions.¹⁹ Manos et al.²⁰ suggest that HPV DNA testing may help identify underlying HSILs in women with ASCUS Pap smears. Analysis of data from natural history studies, cross-sectional data, the ALTS trial, and clinical trials⁵ may help guide future modifications to the Bethesda System and possibly help clarify the term ASCUS.

REFERENCES

1. National Cancer Institute Workshop. The 1988 Bethesda System for reporting cervical/vaginal cytologic diagnoses. *JAMA* 1989;262:931-4.
2. Bottles K, Reiter RC, Steiner AL, Zaleski S, Bedrossian CWM, Johnson SR. Problems encountered with the Bethesda System: the University of Iowa experience. *Obstet Gynecol* 1991;78:410-4.
3. The Bethesda Committee. The Bethesda System for reporting cervical/vaginal cytologic diagnoses. *Acta Cytol* 1993;37:115-24.
4. Kurman RJ, Henson DE, Herbst AL, Noller KL, Schiffman MH. Interim guidelines for management of abnormal cervical cytology. *JAMA* 1994;271:1866-9.
5. Koutsky LA for The ASCUS LSIL Triage Study Group. HPV testing as a triage strategy for women with cytologic evidence of low-grade squamous intraepithelial lesions: early results from a randomized trial on the clinical management of ASCUS and LSIL cytologic abnormalities. *J Natl Cancer Inst* (in press).
6. Kurman RJ, Solomon D. The Bethesda System for reporting cervical/vaginal cytologic diagnoses. New York: Springer-Verlag, 1994:30-43.
7. Schiffman MH, Bauer HM, Hoover RN, Glass AG, Cadell DM, Rush BB, et al. Epidemiologic evidence showing that human papillomavirus infection causes most cervical intraepithelial neoplasia. *J Natl Cancer Inst* 1993;85:958-64.
8. Liaw KL, Glass AG, Manos MM, Greer C, Scott DR, Sherman ME. Detection of human papillomavirus DNA in cytologically normal women and subsequent cervical squamous intraepithelial lesions. *J Natl Cancer Inst* 1999;91:954-60.
9. Sherman ME, Schiffman MH, Lorincz AT, Manos MM, Scott DR, Kurman RJ, et al. Toward objective quality assurance in cervical cytopathology. Correlation of cytopathologic diagnoses with detection of high-risk human papillomavirus types. *Am J Clin Pathol* 1994;102:182-7.
10. Schiffman MH, Kiviat NB, Burk RD, Shah KV, Daniel RW, Lewis R, et al. Accuracy and interlaboratory reliability of human papillomavirus DNA testing by hybrid capture. *J Clin Microsc* 1995;33:545-50.
11. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* 1968;70:213-20.
12. Schiffman MH, Brinton LA. Epidemiology. In: Kurman RJ, editor. Blaustein's pathology of the female genital tract, 4th ed. New York: Springer-Verlag, 1994:1199-223.
13. Robb JA. The "ASCUS" swamp. *Diagn Cytopathol* 1994;11:319-20.
14. Renshaw AA, Lee KR, Granter SR. Use of statistical analysis of cytologic interpretation to determine the causes of interobserver disagreement and in quality improvement. *Cancer (Cancer Cytopathol)* 1997;81:212-9.
15. Davey DD, Naryshkin S, Nielsen ML, Kline TS. Atypical squamous cells of undetermined significance: interlaboratory comparison and quality assurance monitors. *Diagn Cytopathol* 1994;11:390-6.
16. Howell LP, Davis RL. Follow-up of Papanicolaou smears diagnosed as atypical squamous cells of undetermined significance. *Diagn Cytopathol* 1996;14:20-4.
17. Williams ML, Rimm DL, Pedigo MA, Frable WJ. Atypical squamous cells of undetermined significance: correlative histologic and follow-up studies from an academic medical center. *Diagn Cytopathol* 1997;16:1-7.
18. Sherman ME, Kurman RJ. The role of exfoliative cytology and histopathology in screening and triage. *Obstet Gynecol Clinics N Am* 1996;23:641-55.
19. Schiffman MH, Solomon D, Liaw K, Sherman ME. Why, how, and when the cytological diagnosis of ASCUS should be eliminated. *J Lower Gen Tract Dis* 1998;2:165-9.
20. Manos MM, Kinney WK, Hurley LB, Sherman ME, Shieh-Ngai J, Kurman RJ, et al. Identifying women with cervical neoplasia using human papillomavirus DNA testing for equivocal Papanicolaou results. *JAMA* 1999;281:1605-10.