## Title

False-Positive Papanicolaou (PAP) Test Rates in the College of American Pathologists PAP Education and PAP Proficiency Test Programs: Evaluation of False-Positive Responses of High-Grade Squamous Intraepithelial Lesion or Cancer to a Negative Referen...



## Journal

Archives of Pathology & Laboratory Medicine, 138(5)

## ISSN

0003-9985

## Authors

Crothers, Barbara A
Booth, Christine Noga
Darragh, Teresa Marie
et al.

## Publication Date

2014-05-01

## DOI

10.5858/arpa.2013-0083-cp

Peer reviewed

# False-Positive Papanicolaou (PAP) Test Rates in the College of American Pathologists PAP Education and PAP Proficiency Test Programs

## Evaluation of False-Positive Responses of High-Grade Squamous Intraepithelial Lesion or Cancer to a Negative Reference Diagnosis

*Barbara A. Crothers, DO; Christine Noga Booth, MD; Teresa Marie Darragh, MD; Chengquan Zhao, MD; Rhona J. Souers, MS; Nicole Thomas CT(ASCP), MPH; Ann T. Moriarty, MD*

• ***Context.***—**In cytology proficiency testing (PT), participants fail for incorrectly interpreting a high-grade squamous intraepithelial lesion or cancer (HSIL+) Papanicolaou test result as negative. This penalty may lead to a false-positive interpretation of negative slides as HSIL+ to avoid failure.**

***Objective.***—**To investigate factors related to false-positive responses in a PT versus an educational environment.**

***Design.***—**We analyzed 420 079 responses from 9414 validated negative reference slides in the College of American Pathologists Interlaboratory Comparison Program in Gynecologic Cytopathology (PAP Education) and compared them with responses from the Gynecologic Cytology Proficiency Testing Program for the percentage of false-positive (HSIL+) interpretations in each of 7 negative subcategories. We evaluated the influence of preparation type (ThinPrep, SurePath, and conventional Papanicolaou test), participant type (pathologist or cytotechnologist), and program time interval (preproficiency test or PT) on a false-positive response.**

***Results.***—**Reference diagnosis and participant type, but not preparation type, were statistically correlated to false-positive responses. The interaction between program time interval and participant type was also significant. Pathologists had higher rates of false-positive results on preproficiency test (1.2% [800 of 68 690]) than they did on PT (0.8% [993 of 129 857]). Cytotechnologists had no differences between program time intervals (preproficiency, 0.9% [515 of 63 281] versus PT, 1.0 [1231 of 121 621]; $P = .91$). Negative subcategories frequently mistaken for HSIL+ were reparative changes (4.7% [427 of 9069]), atrophic vaginitis (1.8% [18 of 987]), and negative for intraepithelial lesion or malignancy (1.2% [2143 of 178 651]), but during PT, false-positive rates were significantly increased only for the negative for intraepithelial lesion or malignancy and herpes simplex virus ($P < .001$).**

***Conclusions.***—**Pathologists had lower false-positive rates in the Gynecologic Cytology Proficiency Testing Program than they did in PAP Education, but participants were more likely to report a false-positive response (HSIL+) for negative for intraepithelial lesion or malignancy and herpes simplex virus in the Gynecologic Cytology Proficiency Test Program.**

**(*Arch Pathol Lab Med.* 2014;138:613–619; doi: 10.5858/arpa.2013-0083-CP)**

The advent of vaccines against several high-risk subtypes of human papillomavirus should decrease the prevalence of squamous cell carcinoma of the cervix and its precursor, high-grade squamous intraepithelial lesion (HSIL). In addition, recently updated national guidelines have modified cervical cancer screening recommendations to exclude women younger than 21 years and to increase the screening interval to 3 years in women 21 to 29 years and to 5 years in women 30 years or older with negative test results and histories.[1,2] In most pathology practices, the total annual HSIL diagnoses comprise less than 0.8% (90th percentile of laboratories from 2009 participants) of their total Papanico-

Accepted for publication June 10, 2013.

From the Department of Pathology and Laboratory Medicine, Walter Reed National Military Medical Center, Bethesda, Maryland (Dr Crothers); the Department of Anatomic Pathology, Cleveland Clinic Foundation, Cleveland, Ohio (Dr Booth); the Pathology Cytopathology Laboratory, Mount Zion Medical Center Clinic, University of California, San Francisco (Dr Darragh); Department of Pathology, Magee-Women's Hospital of University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania (Dr Zhao); and the Departments of Statistics/Biostatistics (Ms Souers), CAP PAP Program (Ms Thomas), and Pathology (Dr Moriarty), College of American Pathologists, Northfield, Illinois.

Dr Darragh reported receiving research support (supplies) from Hologic, Inc. The other authors have no relevant financial interest in the products or companies described in this article.

The opinions or assertions contained herein are the private views of the authors and do not reflect the official policy of the Department of the Army, Department of Defense, or US government.

Reprints: Barbara A Crothers, DO, Department of Pathology and Laboratory Medicine, Walter Reed National Military Medical Center, 8901 Wisconsin Ave, Bethesda, MD 20889-5600 (e-mail: Barbara.a.crothers.mil@health.mil).

| Table 1. Interpretive Menu for the Papanicolaou Proficiency Test Program | | |
|---|---|---|
| **Category Selection** | **Code** | **Interpretation** |
| Category A | 001 | Unsatisfactory for evaluation |
| Category B | 101 | Negative for intraepithelial lesion or malignancy, not otherwise specified |
| | 111 | Fungal organisms consistent with *Candida* spp |
| | 113 | *Trichomonas vaginalis* |
| | 115 | Cellular changes consistent with herpes simplex virus |
| | 120 | Reparative changes |
| Category C | 201 | Low-grade squamous intraepithelial lesion |
| Category D | 211 | High-grade squamous intraepithelial lesion |
| | 221 | Squamous cell carcinoma |
| | 225 | Adenocarcinoma |
| | 226 | High-grade squamous intraepithelial lesion, carcinoma, and/or carcinoma, not otherwise specified |

laou (Pap) test volume.[3] As the new screening guidelines are adopted and vaccine usage increases, cytologists will be presented with fewer opportunities to correctly identify these lesions. These changes in clinical practice have the potential to exacerbate a cytologist's tendency to "overcall" Pap tests to avoid missing a serious lesion when cellular changes are present that mimic HSIL. Statistically, a Pap test reported as HSIL will be more likely to be a false-positive result.

In the Gynecologic Cytology Proficiency Test Program (PAP PT), sponsored by the College of American Pathologists (CAP), there is a small but consistent rate of negative slides identified by proficiency test (PT) participants as HSIL or greater (HSIL+). In 2008, we initially observed[4] that PT participants reported 0.98% (13 of 1321) of conventional smear, 1.66% (197 of 11 861) of ThinPrep (Hologic, Marlborough, Massachusetts), and 1.59% (46 of 2899) of SurePath (Becton Dickinson and Company, Franklin Lakes, New Jersey) Pap test responses with a reference diagnosis of negative for intraepithelial lesion or malignancy (NILM) as HSIL+. If extrapolated to the general population, a relatively high number of women might receive Pap test results leading to unnecessary intervention (colposcopy and possible biopsy). Our curiosity was piqued. Are there slide variables that influence a participant's decision to "upgrade" a negative slide to an abnormal category? Do pathologists tend to report false-positive results more often than cytotechnologists, or vice versa? If differences in responses are apparent, are they due to the artificial testing environment created by PT? We investigated variables associated with field-validated negative Pap tests slides overcalled as HSIL+ to identify factors that resulted in a false-positive test and compared those results for pathologists and cytotechnologists before and after the initiation of PT.

## MATERIALS AND METHODS

We retrospectively evaluated 420 079 responses from 9414 field-validated negative reference Pap test slide results from the 100 series (NILM, benign reactive or reparative changes, and organisms including herpes simplex virus (HSV), *Trichomonas vaginalis*, and *Candida* spp) of the CAP Interlaboratory Comparison Program in Gynecologic Cytopathology (PAP Education) between 2000 and 2005, before PT, and from the negative reference category (category B) of the PAP PT program between 2006 and 2011. Table 1 summarizes the exact reference diagnoses from the 100 series as they relate to category B in PAP PT. The PAP Education program consisted of 4 quarterly mailings of 5 Pap test slides. The PAP PT program consists of 2 educational mailings of 5 Pap test slides and a 10-slide examination. The examination and educational slide sets reflect the preparation types examined by the laboratory in practice. From 2000 to 2005, all slides were evaluated by participants as part of a Pap Education program intended to expose participants to a variety of examples of gynecologic cytology interpretations. From 2006 to 2011, all of the slides were evaluated by participants as part of the PAP PT program where participants are under pressure to accurately interpret slides as part of their examination to continue to practice gynecologic cytology. The three different types of Pap test slide preparations in the programs were conventional Pap tests, ThinPrep, and SurePath. Data before 2000 were excluded because different slide preparations, such as monolayer preparations, had not been widely implemented, so most of the preparations in circulation were conventional Pap tests. Data from the 2005-D mailing were excluded because they served as a pilot test for PAP PT. From 2006 to 2011, all slides employed in the examination portion of PAP PT had been field-validated from the educational program. Field validation requires program participant concordance of 90% or greater to the reference diagnosis, and the standard error of that percentage must be, at most, 0.05 (SE $<$ 0.05). Field validation generally required 20 or more responses per slide to establish concordance compliant with these guidelines, depending on the diagnosis. The PAP PT examination consisted of 10 Pap tests (of the same preparation type or types evaluated by the individual in practice), with at least one of the following reference diagnoses: unsatisfactory, negative, low-grade squamous intraepithelial lesion (LSIL), and HSIL or cancer (HSIL+). These diagnoses constitute the 4 interpretive categories used for scoring: categories A, B, C, and D, respectively (Table 1). The scoring system implemented by the Clinical Laboratory Improvement Amendments of 1988 (CLIA)[5] mandates that each individual must attain a passing score of 90 or greater on the PT. Pathologists and cytotechnologists have slightly different scoring systems (Table 2). Each slide has a weighted value of 10 for a correct response, but a negative response for a HSIL+ reference diagnosis by a pathologist results in automatic failure despite correct responses for all other slides because it is scored as –5. Cytotechnologists, although having their tests scored slightly differently, still incur the greatest penalty and automatic failure for a negative response on a HSIL+ slide.

Three board-certified pathologists and/or cytopathologists and one cytotechnologist from the CAP Cytopathology Committee independently confirmed the reference diagnosis of all slides before slide inclusion in the PAP Education program (reference validation). Laboratories submitting slides with a reference diagnosis of LSIL, HSIL, or cancer were required to obtain biopsy confirmation of the Pap test diagnosis before slide submission in accordance with regulations established in CLIA. In 2006, CAP offered a national Gynecologic Cytology Proficiency Test Program (PAP PT) using some of the slides from the PAP Education program. These slides had attained 90% or greater interpretive concordance to the reference diagnosis through field validation. CAP staff cytotechnologists reviewed field-validated slides selected for inclusion into the PAP PT program and excluded slides with physical deficiencies (eg, cracks or chips), processing problems (eg, bubbles), or faded stains. Acceptable slides were packaged as PAP PT and circulated among participants under conditions defined for testing. These conditions included the following: (1) participants have 2 hours to complete the

| Table 2. Papanicolaou Proficiency Testing Scoring Systems by Participant Type | | | | |
|---|---|---|---|---|
| | Examinee Response | | | |
| Correct Response Category | Category A, Unsatisfactory | Category B, Negative | Category C, LSIL | Category D, HSIL+ |
| Pathologist 10-slide test | | | | |
| Category A, Unsatisfactory | 10 | 0 | 0[a] | 0 |
| Category B, Negative | 5 | 10 | 0[a] | 0 |
| Category C, LSIL | 5 | 0 | 10 | 5[a] |
| Category D, HSIL+ | 0 | **−5**[b] | 5[a] | 10 |
| Cytotechnologist 10-slide test | | | | |
| Category A, Unsatisfactory | 10 | 0 | 5[a] | 0 |
| Category B, Negative | 5 | 10 | 5[a] | 0 |
| Category C, LSIL | 5 | 0 | 10 | 10[a] |
| Category D, HSIL+ | 0 | **−5**[b] | 10[a] | 10 |

Abbreviations: HSIL, high-grade squamous intraepithelial lesion; LSIL, low-grade squamous intraepithelial lesion.

[a] Scoring for cytotechnologist and pathologist differ.
[b] Bolded numbers signify an automatic failure.

test, (2) primary screening cytologists (cytotechnologists and primary screening pathologists who perform initial Pap test screening without the assistance of a cytotechnologist in their practice) must evaluate the entire slide set independently without abnormal areas marked by another participant, (3) pathologists who are not primary screeners (secondary screening pathologists) are allowed to receive the slide set and interpretations from a participant cytotechnologist who has marked the most abnormal areas, (4) participants may not discuss answers or review slides together during the test, and (5) participants may not consult reference resources or Web sites for assistance with the examination.

Our analysis examined the factors associated with the false-positive rate of slides with a reference diagnosis of negative (category B) and overcalled as HSIL+ (category D). We used a nonlinear mixed model fitted with preparation type, participant type, interval timing (preproficiency testing [pre-PT] and PT), and reference diagnosis, as well as the interaction term between timing and the 3 other factors (preparation type, participant type, and reference diagnosis). The multilevel model included a repeated-measures component to model the slide-factor correlation structure because repeated measures (responses) are collected on specific individual slides. This model allows us to examine differences in the pattern of responses over time, controlling for multiple responses per slide. We used a significance level of .05 for this analysis. We compared the results for the slides directly, without correcting for repeated measures, and there were no significant changes in the results. We calculated the false-positive rate for the 3 preparation types, 3 participant types (pathologist, cytotechnologist, and the laboratory as a whole), and time interval (pre-PT and PT) (Table 3). We also calculated the false-positive rate (for a category D, HSIL+, response to a category B, negative reference diagnosis) for the 3 participant types, for both pre-PT and PT, and for each of the specific reference diagnoses in the negative category (100 series): NILM, not otherwise specified; fungal organisms consistent with *Candida* spp; *Trichomonas vaginalis*; cellular changes consistent with HSV; reparative changes; atrophic vaginitis; and follicular cervicitis (Table 4). After calculating those results, we tested for performance differences between individual participant types (pathologist versus cytotechnologist) for the 2 diagnoses that had significant performance differences by time interval.

## RESULTS

Table 3 shows the factors associated with a false-positive response from the combined PAP Education and PAP PT programs. The type of slide preparation (conventional, ThinPrep, or SurePath) had no effect on the false-positive rate ($P = .34$) regardless of the reference diagnosis. The participant type and reference diagnosis were significantly associated with the false-positive rate ($P < .001$). The laboratory had significantly fewer false-positive responses (0.7%; 254 of 36 630) in the pre-PT period than did

| Table 3. Factors Associated With a False-Positive Response in the 2000–2011 College of American Pathologists Interlaboratory Comparison Program in Gynecologic Cytopathology (Pre-PT) and the Gynecologic Cytology Proficiency Test Program (PAP PT) | | | |
|---|---|---|---|
| Factor | Responses, No. | False-Positive Rate, No. (%) | *P* Value |
| Preparation type | | | .34 |
| Conventional | 171 147 | 1757 (1.0) | |
| ThinPrep | 192 308 | 1718 (0.9) | |
| SurePath | 56 624 | 378 (0.7) | |
| Participant type | | | <.001 |
| Pathologist | 198 547 | 1793 (0.9) | |
| Cytotechnologist | 184 902 | 1806 (1.0) | |
| Laboratory | 36 630 | 254 (0.7) | |
| Time interval | | | .54 |
| Pre-PT (2000–2005) | 168 601 | 1629 (1.0) | |
| PAP PT (2006–2011) | 251 478 | 2224 (0.9) | |
| Negative reference diagnosis | | | <.001 |
| Negative for intraepithelial lesion or malignancy | 178 651 | 2143 (1.2) | |
| Fungal organisms consistent with *Candida spp*. | 101 157 | 396 (0.4) | |
| *Trichomonas vaginalis* | 99 015 | 626 (0.6) | |
| Cellular changes consistent with herpes simplex virus | 31 184 | 243 (0.8) | |
| Reparative changes | 9069 | 427 (4.7) | |
| Atrophic vaginitis | 987 | 18 (1.8) | |
| Follicular cervicitis | 16 | 0 (0.0) | |

**Table 4. Effect of Participant Type and Reference Diagnosis on False-Positive Response Rates During the Time Interval of the College of American Pathologists Interlaboratory Comparison Program in Gynecologic Cytopathology (pre-PT) and the Gynecologic Cytology Proficiency Test (PT) Programs**

| Factor[a,b] | Responses, No. | False-Positive Rate, No. (%) | *P* Value |
|---|---|---|---|
| Participant type by time interval | | | <.001 |
| Pathologist | | | |
| Pre-PT | 68 690 | 800 (1.2) | <.001 |
| PT | 129 857 | 993 (0.8) | |
| Cytotechnologist | | | .91 |
| Pre-PT | 63 281 | 575 (0.9) | |
| PT | 121 621 | 1231 (1.0) | |
| Laboratory | | | |
| Pre-PT | 36 630 | 254 (0.7) | |
| PT | 0 | | |
| Reference diagnosis by time interval | | | .34 |
| Negative for intraepithelial lesion or malignancy | | | |
| Pre-PT | 84 191 | 845 (1.0) | <.001 |
| PT | 94 460 | 1298 (1.4) | |
| Fungal organisms consistent with *Candida* species | | | .11 |
| Pre-PT | 24 450 | 101 (0.4) | |
| PT | 76 707 | 295 (0.4) | |
| *Trichomonas vaginalis* | | | .65 |
| Pre-PT | 42 481 | 247 (0.6) | |
| PT | 56 534 | 379 (0.7) | |
| Cellular changes consistent with herpes simplex virus | | | .04 |
| Pre-PT | 8494 | 49 (0.6) | |
| PT | 22 690 | 194 (0.9) | |
| Reparative changes | | | .56 |
| Pre-PT | 7987 | 369 (4.6) | |
| PT | 1082 | 58 (5.4) | |
| Atrophic vaginitis | | | |
| Pre-PT | 982 | 18 (1.8) | No test |
| PT | 5 | 0 (0.0) | |
| Follicular cervicitis | | | |
| Pre-PT | 16 | 0 (0.0) | No test |
| PT | 0 | NA | |

Abbreviation: NA, not applicable.

[a] Pre-PT includes responses from the 2000–2005 College of American Pathologists Interlaboratory Comparison Program in Gynecologic Cytopathology.

[b] PT includes responses from the 2006–2011 College of American Pathologists Gynecologic Cytology Proficiency Test Program.

pathologists (0.9%; 1793 of 198 547) or cytotechnologists (1.0%; 1806 of 184 902), but there were no significant differences between false-positive responses for pathologists and cytotechnologists. There was a significant difference in the false-positive response rate related to specific subcategories in a negative reference diagnosis ($P < .001$). Overall, most false-negative responses occurred with an exact reference diagnosis of reparative changes (4.7%; 427 of 9069), atrophic vaginitis (1.8%; 18 of 987), NILM (1.2%; 2143 of 178 651), and cellular changes consistent with HSV (0.8%; 243 of 31 184).

Table 4 shows the interaction between participant type (pathologist, cytotechnologist, or laboratory) and exact reference diagnosis in the negative category during the time interval (pre-PT versus PT). The interaction between the time interval of the slide challenge and the participant type was significant ($P < .001$), showing that pathologists had higher false-positive rates for pre-PT (1.2%; 800 of 68 690) as opposed to the PT time interval (0.8%; 993 of 129 857). Compared with pathologists, cytotechnologists had lower false-positive rates for pre-PT (0.9%; 575 of 63 281), but their PT false-positive rate (1.0%; 1231 of 121 621) was not significantly different between time intervals ($P = .91$). There is no PT result for the laboratory as a whole because only individuals are graded in PAP PT. The interaction between exact reference diagnosis by time interval was not significant ($P = .34$). However, the false-positive rates were significantly different between time intervals for the diagnosis of NILM ($P < .001$) and for cellular changes consistent with HSV ($P = .04$). We tested separately for performance differences between pathologists and cytotechnologists for these 2 reference diagnoses in the negative category because they had significant performance differences. The overall false-positive rate for NILM challenges in both programs was 1.2% for pathologists (957 of 82 822) and 1.4% for cytotechnologists (1046 of 77 365) ($P = .09$; results not in table). For cellular changes consistent with HSV challenges, the pathologist and cytotechnologist's false-positive rates were 0.7% (111 of 15 359) and 0.9% (123 of 14 024), respectively, but this was not statistically significant ($P = .73$). Follicular cervicitis cases were not included in the PAP PT, and few ($n = 5$) atrophic vaginitis cases were included because it has been difficult to obtain 90% concordance on these slides in PAP Education.

## COMMENT

From 2006 to 2011, all of the slides in our study had been field-validated from the PAP Education program, but in PAP PT examination challenges, participant conditions were different. To comply with CLIA, participants are required to pass one national gynecologic cytology proficiency examination annually or cease interpretation of Pap tests.[5] The

scoring system required by CLIA penalizes participants for a *false-negative* interpretation because each individual must attain a passing score of 90. Because participants gain points for each correct or partially correct answer, those who interpret a HSIL+ slide as negative will never attain a passing score, even if all other 9 responses are correct, because that mistake deducts points from their total score (9 × 10 points = 90 points; 90 − 5 points = 85 points ). Moriarty et al[6] reported that 1% of HSIL+ slides in 2006 to 2007 PAP PT were reported as negative by participants, and that this was lower than the false-negative rate that was projected (1.2% for conventional slides and 2.2% for liquid-based preparations) from the 2004 PAP Education program data. They deduced that one of the factors influencing the low false-negative rate was the participant's likelihood to "game" the system in PAP PT. In other words, the participant would choose HSIL+ as the response if the differential diagnosis for the slide was between negative and HSIL+. Other factors that might have contributed to the lower false-negative rate included removing poorly performing slides from the program in 2007, the elimination of individuals who failed initial PAP PT in 2006, and the difference in the slide review environment for the 2 periods: PAP Education, whereby participants are free to make errors and learn from them, and PAP PT, where participants are penalized for error. Using the data from these programs, in this study, we wanted to investigate the potential reasons for a false-positive (HSIL+) response.

We hypothesized that during PAP PT, participants may try to game the system to prevent automatic failure, and when the differential diagnosis rests between negative and HSIL+, would err on the side of overcalling the slide HSIL+, a false-positive result. Because cytotechnologists are scored slightly less severely, we wondered if they would also err on the side of a false-positive diagnosis in a PT setting. Hughes et al[7] studied the error rate for participants in the 2006 PAP PT (the first year that the CAP offered PT) and compared them with historic error rates on the same slides in PAP Education and discovered that slide performance was different in PAP PT. Both pathologists ($P < .002$) and cytotechnologists ($P = .001$) were more likely to report a negative slide (category B) as abnormal (category C or D) in PAP PT, indicating that a "defensive strategy" might have been employed to ensure success in the testing environment.[7]

For the time interval of 2000–2005 (pre-PT), most participants would not have been professionally penalized for a discordant interpretation. During that interval, some institutions may have used the PAP Education program as one of several forms of individual performance evaluation, but the program was designed to compare laboratories, not individuals. Cytology professionals generally acknowledge the value of cooperative interaction for the interpretation of difficult cytology slides and educational programs encourage that approach, but consultation with one's colleagues is not permitted during PAP PT. Our findings indicate that participants are more likely to report a false-positive HSIL+ in the PAP PT program as opposed to the PAP Education program, but only for those slides in the negative category with a reference diagnosis of NILM or cellular changes consistent with HSV. These data suggest that slides with the reference diagnosis of NILM or HSV display characteristics that mimic HSIL+, causing the participants to have to make a choice between a negative result and a significantly abnormal result. Otherwise, there is no statistically significant difference between false-positive rates in the negative category between the pre-PT and PT testing periods. The interaction between timing and reference diagnosis is not significant because the direction of the performance per diagnosis is the same, except for the diagnosis of fungal organisms consistent with *Candida*. As shown in Table 4, the false-positive rate increased slightly for most negative reference diagnoses during PT. For the diagnosis of fungal organisms consistent with *Candida* spp, there was no increase in false-positive rates; pre-PT and PT rates were equal (0.4%; 101 of 24 450 and 295 of 76 707, respectively). This means that participants were just as likely to mistakenly respond with one of these reference diagnoses in pre-PT and in PT, so gamesmanship was not likely the reason for error. When we tested for performance differences for participant type (pathologist versus cytotechnologist) for NILM and cellular changes consistent with HSV false-positive responses, we found that pathologists had a slightly lower false-positive rate for both responses (1.2% [957 of 82 822] and 0.7% [111 of 15 359], respectively) than did cytotechnologists (1.4% [1046 of 77 365] and 0.9% [123 of 14 024], respectively), but it was not statistically significant. Thus, both are equally likely to misinterpret an NILM or cellular changes consistent with HSV slide as HSIL+ in PT. Slides showing early herpetic changes might understandably be called HSIL+ because single, infected cells display hyperchromasia, chromatin clumping, nuclear membrane irregularities, and a high nuclear to cytoplasmic ratio, before developing the characteristic features of multinucleation, nuclear molding, ground-glass chromatin, and chromatin margination. Moriarty et al[8] have reported previously on the performance of herpes simplex slides in proficiency testing. In that study, they determined that changes in slide performance at pre-PT and PT were likely due to the testing environment, as opposed to difficulty interpreting herpes changes on slides. There was no significant difference in HSV slide performance during educational challenges: the only significant difference occurred during PT. These findings are mirrored in our study. Participants vacillating between a negative interpretation of HSV and a HSIL+ response during PT might answer with a false-positive result to avoid failure. Because both participant types (pathologists and cytotechnologists) called HSV false-positive equally often, one wonders if the error was in screening as opposed to interpretation. After all, herpes has distinctive cytologic features once diagnostic cells are located, but on any given slide, diagnostic cells may be infrequent. If the secondary pathologists in PAP PT relied on the primary cytotechnologists to screen and locate diagnostic cells, and the cytotechnologists failed to identify rare diagnostic HSV cells, then both participants might make the same interpretive error.

What remains unanswered by our study is, what features of NILM slides prompted a HSIL+ diagnosis? Review of the slides that were erroneously reported as false-positive was beyond the scope of this study, so our comments on the reasons for false-positive results in NILM are speculative. Could the performance of these slides be due to processing as opposed to interpretive problems? We chose a multilevel model with a repeated-measures component for analysis of responses to slides to correct for multiple views of the same slides but later applied a basic comparison model that compared the results directly. The results were the same with both models. If the model results had been different, a subset of slides might have been weighing the performance

toward false-positive results, which would mean that modeling correlation structure would be necessary, and that was not the case. This suggests that multiple views of the slides were not critical to our results. In other words, the slides were not simply getting worse over time because of fading stains or other degradation that might influence the participant's responses.

*Negative for intraepithelial lesion or malignancy* is a broad category that encompasses many normal cellular changes on Pap test slides. Each patient's slide is different, and there are many varieties of "normal" that include minor reactive changes. Normal slides also contain variable numbers of cell types other than squamous cells, including glandular cells, metaplastic cells, and inflammatory cells, all of which might serve as foils for abnormal cells. Any accepted, negative slide that was not otherwise submitted under a specific descriptive diagnosis in category B (Table 1) would have been designated as NILM and could include many mimics of other lesions. For instance, some slides with atrophic changes, but without inflammation, might have been included as NILM, and that might account for a higher false-positive rate in the NILM subcategory. According to Crothers et al, atrophic squamous cells can mimic HSIL+, and atrophic vaginitis slides are designated as HSIL+ in the Pap Education program by up to 1.9% of participants, depending on preparation type.[9] Atrophic vaginitis slides do not perform well in the PAP Education program, possibly because definitive criteria for atrophic vaginitis are vague. We have previously identified that the changes best recognized as atrophic vaginitis (degenerating parabasal cells, necrotic background, and pseudoparakeratotic cells) are the changes of severe atrophy and that inflammation, although important, is not the best or only discriminator.[9]

Cytotechnologists tend to categorize slides with a reference diagnosis of atrophic vaginitis as NILM more often than pathologists do, but both are equally likely to call these cases HSIL+.[9] Endometrial and endocervical cells can also mimic HSIL and can be included in NILM.

Our study indicates that the negative subcategories most frequently mistaken for HSIL+ were reparative changes (pre-PT, 4.6% [369 of 7987]; PT, 5.4% [58 of 1082]), atrophic vaginitis (pre-PT, 1.8% [18 of 982]; no cases for PT [0 of 5]), and NILM (pre-PT, 1.0% [845 of 84 191]; PT, 1.4% [1298 of 94 460]). Reparative changes have been previously identified by the Cytopathology Committee as causing difficulty in the PAP Education program. Colgan et al[10] studied the performance of conventional Pap tests with reparative changes in the 1998 CAP Interlaboratory Comparison Program in Cervicovaginal Cytology and found reparative changes yielded the most false-positive laboratory responses in the negative category and was also a significant cause of false-negative responses. Snyder et al[11] studied the performance of ThinPrep slides with reparative changes using slides from the 2000 to 2003 CAP Interlaboratory Comparison Program in Cervicovaginal Cytology. Similar to Colgan et al,[10] they found reparative changes resulted in the most false-positive responses in the negative category, and that individual participants (both cytotechnologists and pathologists) were more likely to make that error than the laboratory as a whole. Overall, the rate of false-positive interpretations from a slide with reparative changes was significantly less for ThinPrep slides (7.1%) when compared with conventional smears (15.7%; $P < .001$). Our results corroborate those findings but also indicate that these diagnoses (atrophic vaginitis and reparative changes) are equally difficult in PAP PT, and the PT environment is not the primary reason for a false-positive response. The false-positive rates for NILM and HSV challenges are significantly different between pre-PT and PT, which suggests that participants are overcalling these negative challenges as HSIL+ when the differential diagnosis includes HSIL+.

In our study, there was no statistically significant performance difference for reporting a false-positive diagnosis between pathologists and cytotechnologists for PAP Education and PAP PT for 2000 to 2011, even though both participant types had significantly higher rates of false-positive results ($P < .001$) than did the laboratory. This finding supports the practice of sharing difficult cases with one's colleagues because a majority opinion is more likely to result in a correct interpretation, leading to more-appropriate patient care.

There was a significant difference ($P < .001$) between pathologist and cytotechnologist performance between the time intervals (pre-PT and PT). Pathologists were significantly less likely ($P < .001$) to make a false-positive diagnosis during PT (0.8%) than they were in the PAP Education program (1.2%), the opposite of what we had hypothesized. Cytotechnologists showed no significant improvement in performance for false-positive responses for the 2 time intervals. Overall, there was no difference between the false-positive rates for the pre-PT and PT time intervals ($P = .54$). Our statistical model accounts for the increased numbers of responses in each category for PAP PT. A recent study examining the first 4 years of data from CAP PAP PT demonstrated that primary (screening) pathologists fail PAP PT mostly because of *false-positive* responses, whereas secondary pathologists and cytotechnologists have more automatic failures from *false-negative* responses.[12] This study also demonstrated that cytotechnologists are least likely to fail PAP PT and that primary screening pathologists are most likely to fail. Secondary pathologists may be less likely to make a false-positive diagnosis in PAP PT because they receive their test slides prescreened and marked by a cytotechnologist who is also taking the test. Secondary pathologists are also permitted to review the cytotechnologist's interpretations for test slides. In PAP Education, pathologists may simply screen the slides and review them without the benefit of the cytotechnologist's opinion because participants may review the slides independent of other participants. Secondary pathologists may be accepting the cytotechnologist's answers as the correct response in PAP PT, even if they initially disagree with the cytotechnologist's interpretation. This might explain why primary screening pathologists, who do not have the benefit of a cytotechnologist's interpretation, are more likely to fail because of false-positive responses. Pathologists had lower false-positive rates in PAP PT than in PAP Education for 3 interpretations—fungal organisms consistent with *Candida* spp, *Trichomonas vaginalis*, and reparative changes. In practice, most infectious slides are reported by cytotechnologists only and are rarely viewed by pathologists, giving them less exposure to the spectrum of cellular changes seen in these processes. In an educational environment, pathologists may be inclined to "overcall" these changes. In practice, pathologists are generally biased (by virtue of the types of Pap tests triaged to them) toward reporting slides as abnormal and, therefore, may be innately inclined toward higher false-positive rates than are cytotechnologists. In PAP PT, cytotechnologists, who are more experienced with benign processes, might be more likely to

correctly identify minor changes as part of an infectious process. Secondary pathologists who accept the cytotechnologist interpretation in those instances would be less likely to make an interpretive error or report a false-positive result. Because the cytotechnologist's primary responsibility is to identify abnormal slides for pathologists' review, their false-positive rates, not surprisingly, would be unchanged between PAP Education and PAP PT.

The type of slide preparation had no significant effect on a false-positive response ($P = .34$). That may be because participants generally select the preparation type with which they are most familiar. For PAP PT, participants must select the preparation type that they evaluate routinely, but in PAP Education, participants can select other preparations for practice or learning, so if there were a difference in performance of slides, we would have expected to see it during the pre-PT interval. Another reason for the lack of influence of preparation type on the false-positive response rate may be due to extensive field validation for the slides. Participants notify the CAP when slides are poorly preserved, are poorly stained, contain air-bubbles, or have faded, and these are removed from the program. Participants will also often challenge difficult cases, which are reviewed by committee members and may be removed from the program. Subsequently, only the best examples in each category perform well enough to become part of PAP PT. This appears to introduce bias into the PAP PT data because slides must be field-validated with 90% concordance, but we chose the repeated-measures statistical model to control the effect of poorly performing slides with multiple responses in the PAP Education program. Because the false-positive rate is the same for both time intervals, it is unlikely that this bias had a significant effect on the results.

In summary, this study supports the daily practice of involving multiple individuals in the interpretation of Pap tests to prevent false-positive results. Cytotechnologists may serve a protective role in preventing overinterpretation of certain reactive and infectious changes by pathologists as HSIL+ in a testing environment. In the future, pathologists and cytotechnologists will see fewer examples of HSIL+ on Pap tests as a natural result of decreased disease prevalence secondary to human papillomavirus vaccination and decreased Pap volume per laboratory with adherence to new screening guidelines. Human papillomavirus testing has and will continue to replace Pap tests as an alternative test to determine risk of developing cervical cancer in a subset of the population, further reducing pathologists' and cytotechnologists' exposure to HSIL+. This may eventually translate to an increase in false-positive interpretations of HSIL+ on negative Pap tests. Our study shows that participants in PT are likely to overcall some slides (NILM and HSV) because of the test environment. However, certain mimics of HSIL+ (reparative changes, atrophy, HSV, and endometrial cells) continue to challenge participants both in educational and PT environments and continue to contribute to false-positive diagnoses in Pap testing, regardless of interpreter or preparation type.

The study was supported by the College of American Pathologists (CAP), but conceptualized by members of the CAP Cytopathology Committee.

**References**

1. Saslow D, Solomon D, Lawson HW, et al; American Cancer Society; American Society for Colposcopy and Cervical Pathology; American Society for Clinical Pathology. American Cancer Society, American Society for Colposcopy and Cervical Pathology, and American Society for Clinical Pathology screening guidelines for the prevention and early detection of cervical cancer. *Am J Clin Pathol*. 2012;137(4):516–542.

2. Moyer VA; U.S. Preventive Services Task Force. Screening for cervical cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2012;156(12):880–891.

3. College of American Pathologists. Cytopathology checklist: CAP Accreditation Program—Reference, Resources, and Publications page. CAP Web site (behind firewall). http://www.cap.org/apps/cap.portal?_nfpb=true&OnlineChecklistController_8_6_actionOverride=%2Fportlets%2FOnlineChecklist%2FprepareChecklist&_windowLabel=OnlineChecklistController_8_6&_pageLabel=customChecklists_page. Updated July 31, 2012. Accessed December 10, 2012.

4. College of American Pathologists, Cytopathology Committee. PAP Gynecologic Cytology PT program (PAP PT): year end summary report—CAP surveys and educational anatomic pathology programs, 2008. http://www.cap.org/apps/docs/committees/cytopathology/2008_pap_yesr_final.pdf. Accessed October 12, 2012.

5. US Department of Health and Human Services, Health Care Financing Administration. Clinical laboratory improvement amendments of 1998: final rule. *Fed Regist*. 1992;57(40):7146. Codified at 42 CFR §493.1274(c)(2); §493.945.

6. Moriarty AT, Crothers BA, Bentz JS, Souers RJ, Fatheree LA, Wilbur DC. Automatic failure in gynecologic cytology proficiency testing: results from the College of American Pathologists proficiency testing program. *Arch Pathol Lab Med*. 2009;133(11):1757–1760.

7. Hughes JH, Bentz JS, Fatheree L, Souers RJ, Wilbur DC. Changes in participant performance in the "test-taking" environment: observations from the 2006 College of American Pathologists gynecologic cytology proficiency testing program. *Arch Pathol Lab Med*. 2009;133(2):279–282.

8. Moriarty AT, Darragh TM, Souers R, Fatheree LA, Wilbur DC. Performance of herpes simplex challenges in proficiency testing: observations from the College of American Pathologists (CAP) Proficiency testing (PT) Program. *Arch Pathol Lab Med*. 2009;133(9):1379–1382.

9. Crothers BA, Booth CN, Darragh TM, et al. Atrophic vaginitis: concordance and interpretation of slides in the College of American Pathologists cervicovaginal interlaboratory comparison program in gynecologic cytopathology. *Arch Pathol Lab Med*. 2012;136(11):1332–1338.

10. Colgan TJ, Woodhouse SL, Styer PE, Kennedy M, Davey DD. Reparative changes and the false-positive/false-negative Papanicolaou test: a study from the College of American Pathologists Interlaboratory Comparison Program in Cervicovaginal Cytology. *Arch Pathol Lab Med*. 2001;125(1):134–140.

11. Snyder TM, Renshaw AA, Styer PE, Mody DR, Colgan TJ; Cytopathology Resource Committee, College of American Pathologists. Altered recognition of reparative changes in ThinPrep specimens in the College of American Pathologists gynecologic cytology program. *Arch Pathol Lab Med*. 2005;129(7):861–865.

12. Ducatman BS, Bentz JS, Fatheree LA, et al; Cytopathology Resource Committee. Gynecologic cytology proficiency testing failures: what have we learned?: observations from the College of American Pathologists gynecologic cytology proficiency testing program. *Arch Pathol Lab Med*. 2011;135(11):1442–1446.