

Cytopathology and More | Cytologic-histologic correlation: And the answer is...

Barbara A. Crothers, DO

August 2013—You have a great gynecologic cytology case, a patient with atypical endometrial cells on Pap test that you believe might represent a low-grade endometrial adenocarcinoma, but it has been four weeks and you have had no feedback about the patient's outcome. It seems as if there have been a lot of atypical endometrial cells on Pap tests lately. Could it be due to the implementation of a new liquid-based technology for Pap tests in your laboratory? Fortunately, your laboratory performs cytologic-histologic correlation monthly, so you ask the medical director if she has noticed any trends in the rate of atypical glandular cells, and what the corresponding biopsies have shown. To your relief, the patient had a biopsy showing low-grade endometrial carcinoma, and the laboratory statistics have shown only a slight increase in atypical glandular cells since the new technology was implemented. The medical director informs you that she has been recording these data as a special QA project to determine if the increase is due to over-interpretation of reactive glandular cells, because the technology enhances nuclear and cytologic details of glandular cells.

What do other laboratories do with cytologic-histologic correlations, and how does it compare with what your medical director does?

The Centers for Disease Control and Prevention in 2010 awarded the CAP a cooperative agreement to investigate, describe, and outline current quality practices in gynecologic cytology, with a goal of establishing standards for common practices and procedures to allow for accurate benchmarking among laboratories. The entire process required months of planning, coordinating with stakeholders, collecting data through surveys (online and mailed), reviewing the literature, and meeting in working groups. In June 2011, the Gynecologic Cytopathology Quality Consensus Conference (GCQC2) was convened, sponsored by the CAP and with the CDC, American Society for Cytopathology, American Society for Clinical Pathology, and American Society for Cytotechnology as partners, to describe and outline effective quality assurance practices in cytopathology, investigate research evidence of effectiveness of specific practices, and gain group consensus for future practices. The proceedings and outcomes are published in a special section of *Archives of Pathology & Laboratory Medicine*.¹

As part of this effort, a Cytologic-Histologic Correlation Working Group was established to investigate measures used for comparing cytologic specimen reports with surgical biopsy outcomes and to review existing literature for the effectiveness of these practices. Although the review focused on gynecologic cytology specimens, the principles for improving current quality assurance practices apply to all cytologic-histologic correlations (CHC).

Cytologic-histologic correlation is a powerful cytopathology quality assurance tool that may be overlooked and underused. For pathologists with limited experience in cytopathology, it is a great educator and feedback mechanism: You get the answer (a biopsy) to your cytologic impression. There are lessons to be learned from both specimens and information to be gained about processes. The elegance of CHC is that regardless of a cytologist's initial interpretation of a cytology specimen, there is usually a subsequent tissue biopsy that can confirm the result or reveal why the initial interpretation was incorrect. Even though interpretations between observers may vary, the correct interpretation eventually becomes clear. This is in contrast to surgical pathology, where disagreement between individuals over an interpretation may not reveal one correct answer, even though that interpretation is considered the gold standard or "truth" in disease diagnosis. Although this limitation of surgical pathology also affects CHC, in most instances comparing the two specimens is straightforward. Additionally, CHC opens the door for learning experiences in surgical pathology where there are interpretive disagreements about biopsy results.

Cytologic-histologic correlation for gynecologic pathology serves two broad purposes in quality assurance:

- It provides critical information on necessary patient followup by resolving

Pap-biopsy discrepancies or confirming discrepant diagnoses as correct, or both.

- It provides a mechanism with which to monitor the performance and processes of the laboratory to improve overall quality.

The CHC Working Group and consensus conference participants came to several broad conclusions about gynecologic cytology CHC. They are summarized here.

Cytologic-histologic correlation may be performed real time or retrospectively or both.

Real-time correlation implies that available cytology slides are reviewed in conjunction with the surgical biopsy or relatively soon after the surgical biopsy is evaluated, but before a surgical biopsy report is issued. The advantage of concurrent review is that it has a greater impact on immediate patient care. It allows the pathologist to provide health care professionals with critical followup information to the cytology in the surgical report and to resolve or discuss discrepancies between the two specimens, if any, in that report. It is strongly preferred in instances where the Pap test is interpreted as high-grade squamous intraepithelial lesion (HSIL) and the followup cervical biopsy is negative, regardless of the outcome of the review of these specimens.

Retrospective correlation is a review of slides after both reports are issued and serves as a monitor of cytology and biopsy performance and processes for laboratory quality improvement. Data collection is more easily performed retrospectively, since computer software can collaborate findings over time in predesigned reports. Retrospective review is logistically more tenable for laboratories with high volumes or that may not have resources for timely concurrent review. Retrospective review is the only option for laboratories that do not receive biopsies on all patients with cytology results from their laboratory.

Bidirectional correlation, or performing CHC real-time and retrospectively, is the most common practice among laboratories, probably because they serve different purposes. Regardless of the method of correlation, results should be reported in a quality assurance document and monitored over time to identify laboratory trends. Results can be monitored weekly, monthly, quarterly, or annually depending on laboratory volume.

At a minimum, review all available slides for high-grade squamous intraepithelial lesion (HSIL) Pap tests with negative biopsies, with a correlation interval between three to four months but not exceeding six months.

Reviewing both the cytology and surgical biopsies for accuracy of interpretation benefits the patient and the laboratory. Review of an HSIL Pap test that reveals erroneous interpretation of metaplastic cells as HSIL could prevent unnecessary procedures. Confirming the original HSIL Pap test result after a negative biopsy is equally important if the health care professional did not find colposcopic evidence of disease, since the lesion may be hidden in the endocervical canal. If the CHC review occurs in real time, pathologists should take all necessary steps to ensure adequate biopsy orientation and leveling to unveil hidden squamous intraepithelial lesion (SIL).

The preferable retrospective search for a Pap test to correlate on a particular patient with a current biopsy is within three to four months, but no longer than six months, from the time of the biopsy. For laboratories that search their databases for retrospective correlation, most patients will have had an incident Pap test resulting in a biopsy within four months. Reviewing Pap tests older than six months from the time of biopsy could result in false-negative results if the lesion evolved or resolved.

For correlation purposes, the “incident” or first prior Pap test with a significant abnormality should be correlated with the most abnormal current tissue obtained. Pap tests taken in conjunction with a biopsy can be excluded from CHC unless the laboratory has no knowledge of the incident Pap test. Additionally, endocervical curettage without cervical biopsies can be excluded unless they contain a squamous or glandular lesion, and excisional biopsies such as loop electrocautery excisional procedures (LEEP), cervical conization, and hysterectomies should be included.

Laboratories using HSIL or cancer biopsy targets should correlate with the most abnormal prior Pap test taken within the past six months, and exclude Pap tests taken concomitant with the biopsy unless no earlier Pap test is available.

Laboratories may choose to review Pap tests with potentially less clinically significant lesions, such as atypical squamous cells (ASC), low-grade squamous intraepithelial lesion (LSIL), or reactive changes as a part of CHC when multiple slides are available, but HSIL and cancer are the minimal considerations for review because of the significance of these diagnoses.

Standardization of CHC and its metrics is desirable.

Despite decades of reviewing slides for CHC, there remain no acceptable standards for CHC performance or collection of data metrics. As a result, laboratories have no means of comparing their data with those of their peers. One of the charges of the working group was to offer evidence-based standards for data collection that would allow for peer-to-peer comparison. As result of their literature review, the working group proposed that laboratories monitor the following parameters: **1)** the total number of CHC pairs, **2)** the number of positive correlations (“true positives,” as defined prior to actual CHC review of the specimens), **3)** the number of negative correlations (“false-positives,” as defined prior to review), and **4)** the positive predictive value (PPV) of a positive Pap test. These statistics should be tabulated at least annually, although it is appropriate for high-volume laboratories to collect these statistics more frequently.

Most laboratories (78 percent of laboratories responding to the survey) already tabulate these statistics, with the exception of the PPV. As a standard, laboratories should use, at a minimum, the definitive of a true positive and true negative as shown in **Table 1**. If the correlation contains any of the elements from the left and right sides of the table, then it constitutes a positive correlation (abnormality suspected and confirmed). Readers will notice that atypical squamous or glandular cells are not counted in the standard definition of a positive correlation. This does not prevent laboratories from calculating two separate positive predictive values—one with and one without the inclusion of atypical interpretations. Our review of the existing literature on Pap test interpretations of atypical squamous and glandular cells shows very poor inter- and intraobserver concordance, indicating that an atypical Pap test interpretation is not reproducible. This is the primary reason why atypical interpretations were excluded from standardized statistical analysis. A negative correlation is any normal, negative, reactive, or infectious biopsy or Pap test result paired with any interpretation from Table 1.

Pap test result	Cervical biopsy/LEEP/cone/hysterectomy result
Low-grade squamous intraepithelial lesion (LSIL)	HPV changes, cervical intraepithelial lesions 1–3 (CIN1–3)(LSIL, HSIL, or carcinoma in situ (CIS)
High-grade squamous intraepithelial lesion (HSIL)	Dysplasia, unable to grade
Squamous intraepithelial lesion (SIL)	Adenocarcinoma in situ (AIS)
Adenocarcinoma in situ (AIS)	Carcinoma (squamous, glandular, NOS)
Carcinoma (squamous, glandular, NOS)	Other malignant diagnosis
Other malignant diagnosis	

The positive predictive value of a positive Pap test is the preferred standard CHC metric, and laboratories should use the PPV for the whole laboratory to formulate QA monitors.

Evidence shows that the PPV is the most reproducible statistic for CHC.² Most laboratories already collect the data to calculate the PPV but are not aware of the formula to do so. The PPV is defined by the formula:

$$\text{PPV} = \frac{\text{True Positives}}{\text{True Positives} + \text{False-Negatives}}$$

where a true positive is a positive correlation pair and a false-positive is a positive Pap test with a negative biopsy. Notice that the PPV is based on the original interpretation for both the Pap test and the biopsy, and not the review

interpretation of these specimens. The calculation assumes the biopsy is the gold standard of “truth.” The PPV emphasizes the screening role of a Pap test. It is intended to identify women who require triage to colposcopy to confirm a potential abnormality through visual inspection or biopsy or both. To ensure meaningful data, a minimum of 20 total correlation pairs is necessary to calculate PPV.

One reason for the superiority of PPV over metrics such as sensitivity and specificity is that it uses easily retrievable data. Sensitivity and specificity rely on knowing the false-negative (sensitivity=true positives/true positives + false-negatives) or true negative (specificity=true negative/ true negative + false-positive) results. These data are difficult to accurately measure because most women with negative Pap tests are not biopsied. False-positive Pap tests are probably overrepresented because patients are referred for biopsies. The PPV is a measurement that is close to the percent of positive Pap tests that correlate with biopsies. This was the most frequently measured CHC statistic in the laboratory survey. According to CAP Q-Probes data from 2005 to 2010, the median PPV is 83 percent to 88 percent, with a range of 71 percent to 94 percent.² It is important to emphasize that the PPV is a laboratory, not an individual, metric. It would be difficult to obtain an accurate PPV for individuals except in laboratories with a very high volume. Additionally, the PPV does not indicate truth. Review of CHC slides often reveals interpretive or processing errors in both specimens that should not be held against individuals.

If the laboratory’s PPV is low relative to benchmarks, it should investigate Pap interpretive accuracy and intradepartmental variability as part of its QA program. If a laboratory’s PPV is high, it may indicate that the laboratory is identifying only the most obvious lesions and under-recognizing subtle changes. It may also indicate that health care professionals are not sampling subtle colposcopic lesions or are not sampling the transformation zone.

It is desirable to provide timely notification to a caregiver for confirmation of a negative biopsy and HSIL or cancer (HSIL+) Pap test, or of a negative biopsy and an HSIL or cancer Pap test re-interpreted as NILM (negative for intraepithelial lesion or malignancy).

There are significant followup implications for patients with a cytology interpretation of HSIL—most will have an ablative procedure or excisional biopsy. An unintended consequence of cervical cone excisional and LEEP procedures is cervical incompetence. When biopsies are negative, informing the health care provider that a Pap test was correctly interpreted as HSIL or cancer (HSIL+) after a second review enables him or her to proceed with appropriate ablative therapy with confidence. If the Pap test review yields a mistaken interpretation of HSIL+, unnecessary surgery is prevented. In some cases, consensus regarding the initial Pap test interpretation of a high-grade lesion is not achievable and a diagnostic excisional biopsy will be indicated. There was no consensus opinion on the definition of “timely” notification, but notification should occur as soon as is feasible after the microscopic review of both specimens. Discussions with the health care professional should be documented in the biopsy or cytology report or in a separate QA document.

Laboratories should attempt to obtain correlation biopsy information for all patients with an HSIL or cancer Pap test.

It is a challenge for some laboratories to obtain Pap test or biopsy results if they process and interpret only one or the other specimen type, but for correlation purposes, they should attempt to gain biopsy followup information for all patients with an HSIL+ Pap test. This serves two purposes: It ensures that patients with an HSIL+ Pap test obtain appropriate colposcopy, and it allows the laboratory to confirm its accuracy of an HSIL+ interpretation. Requests for followup information may be by a note in the Pap test report, telephone, e-mail, or other means. Laboratories that process both specimen types from the same patient should request followup information from the health care professional if no biopsy or report of colposcopy is documented six months after the incident Pap test. Finally, laboratories should document attempts to obtain followup information, and the method used to request followup should be made part of the written QA program.

Microscopic review of all slides from discordant Pap test/cervical biopsy pairs (as laboratory-defined) is desirable for CHC.

Even though calculation of the PPV does not require microscopic review of Pap test and biopsy pair mismatches, this exercise is the most rewarding and revealing of the entire process, and laboratories should record review findings in a QA document or specimen report. Review of negative Pap slides when a biopsy is interpreted as HSIL may reveal reasons for interpretive error, such as Papanicolaou stains that are too dark for optimal examination of chromatin, processing problems that obstruct diagnostic criteria, or sampling problems that result in incomplete collection of cells or obscuring factors that hinder correct interpretation. It is primarily through this process, and not calculation of PPV, that laboratories will find quality improvement projects that will enhance their performance. If review of all discordant Pap test/cervical biopsy pairs is not possible, the review should focus on HSIL-normal mismatches for both Pap tests and biopsies. It may be futile to review mismatches in LSIL-normal cases because LSIL lesions regress and appear at uncertain intervals and one would expect mismatches that are not the result of interpretive, sampling, or processing errors. However, HSIL is usually a persistent lesion and the ramifications of a mismatched pair are more severe.

If all of the slides in a mismatched pair are not available, those that are available should be reviewed and the original interpretation on unavailable specimens will be assumed to be correct. Laboratories may define their own non-correlation metrics for QA purposes. For example, a laboratory may want to monitor and review all atypical squamous cells, cannot exclude high-grade squamous intraepithelial lesion (ASC-H) and corresponding biopsies to determine the percent of cases with a significant biopsy finding, and then review those Pap tests where the biopsy was interpreted as HSIL+ to determine whether there are features present that would prompt cytologists to interpret those cases as HSIL in the future.

CHC is optimal with a multilayered approach.

Developing a CHC program that meets the laboratory's needs and addresses perceived laboratory problems is an ideal toward which we all strive. A multilayered approach to CHC allows for customization as well as standardization. Laboratories can drill down on particular areas of concern by developing continuous and interval monitors.

One example of a continuous monitor would be the PPV. An interval monitor may target specific pairs for a predetermined time, for example quarterly, to acquire a snapshot of laboratory performance for that indicator. Continuous monitors may be desirable when laboratories experience high personnel turnover, disruptive environments, or other variables such as new instrumentation that can cause a quality drift.

Corrective action for variances can also be creative. The most popular and favored method of investigating and improving interpretive variances among consensus participants was to review slides in a group. Not only does this method encourage discussion and expose all observers to difficult cases, but it can occur in a non-threatening environment where the participants are unaware of the identity of the original interpreters. A group discussion of mismatches and slides encourages uniformity of interpretation, leverages group experience, and allows observers to share diagnostic clues and practices.

Another layer of CHC is to optimize biopsies during review. Studies have shown that biopsy specimens are often the reason for a "false-positive" Pap test result and additional processing may unveil a cervical lesion.³ Reorienting tissue in the block, obtaining additional levels, performing ancillary studies such as p16, and recording the presence or absence of the transformation zone are all methods of optimizing biopsy performance. Providing sampling data to health care professionals who perform colposcopy and biopsy may help improve biopsy sampling. Laboratories can develop trend-based policies to improve internal practice, such as standardizing the number of levels and serial sections on cervical biopsies and endocervical curettage, pinning LEEP and cone specimens flat to optimize embedded sections, and teaching histotechnologists to recognize ectocervix to embed cervical biopsies properly. A laboratory may choose to monitor characteristics of biopsies over time to troubleshoot mismatches in CHC when the Pap tests appear accurate by recording the presence or absence of a transformation zone, biopsy sizes less than 2 mm, colposcopies with only one to two biopsies, poor biopsy orientation, and requests for additional levels.

Pap test interpretation is most often the focus of CHC slide review but other factors are just as guilty of causing error. For example, a laboratory may choose to record the quality of Pap slides in CHC mismatches, including staining and processing irregularities. There may be patient factors that contribute to interpretive error, such as atrophy, obscuring blood or inflammation, infection, or inadequate shedding of abnormal cells. Some patterns of HSIL are notorious for causing interpretive errors—hyperchromatic crowded groups and small individual HSIL cells with bland nuclei.

Curiosity may prompt further CHC investigations. For example, how often does your laboratory have an LSIL Pap test but an HSIL biopsy? Was the Pap test interpreted as LSIL because of few HSIL cells on the slide, or are HSIL cells usually absent? How many ASC-H Pap tests have an HSIL biopsy, and what does review of those Pap tests reveal? Other pairs that might be interesting to monitor to improve laboratory performance are AIS/LSIL, atypical squamous cells of undetermined significance (ASC-US) with a positive test for human papillomavirus (HPV+) and a SIL biopsy, ASC-US with a negative test for HPV and a SIL biopsy, atypical glandular cells (AGC) and subsequent endocervical or endometrial biopsies, and HSIL Pap tests in pregnant or postpartum women. Any of these monitors can be periodic or continuous, depending on other laboratory metrics or conditions.

When reviewing slides for CHC, minimize observer bias. Such bias occurs when the observer tends to believe the result of one test more than the other, or is influenced by the result on one test when reviewing the other. There are several ways to prevent bias. If there is disagreement between the reviewer and the primary cytologist, one can obtain an additional opinion. If retrospective review is performed, all slides can be randomly combined and the reviewer blinded to the original results, with unveiling of the original results only after review. Cytotechnologists can review all of the Pap tests and pathologists review all of the biopsies. For real-time correlation, all mismatches can be triaged to hierarchical peer review, or specific interpretations such as HSIL Pap tests and biopsies may be referred. Finally, all discrepancies can be reviewed together in a consensus conference for a group decision.

Summary

What is the point of CHC if the data are never used or if the primary stakeholders don't have access to the data? Laboratories have a wealth of information at their disposal if they manage it effectively. The CHC Working Group and consensus participants agreed that CHC should not be unnecessarily proscriptive, because laboratories face different problems and need to tailor their approach to CHC to target potential problem areas.

In an ideal, high-quality performance environment, laboratories would receive both cytologic and histologic specimens from the same patient and be able to correlate these results to improve patient outcomes. That is not possible for most laboratories because care is fragmented and they do not usually have control over what specimens they receive. The guidelines suggested in this article are minimum guidelines for CHC that most laboratories can perform and that allow them to compare their performance against national benchmarks compiled from all laboratories.□

References

1. College of American Pathologists Gynecologic Cytopathology Quality Consensus Conference Working Groups 1-5. Special Section—College of American Pathologists Consensus Conference on Gynecologic Quality. *Arch Pathol Lab Med*. 2013; 137(2):158-219.
2. Jones BA, Novis DA. Cervical biopsy-cytology correlation. A College of American Pathologists Q-Probes study of 22439 correlations in 348 laboratories. *Arch Pathol Lab Med*. 1996;120(6):523-531.
3. Bewtra C, Pathan N, Hashish H. Abnormal Pap smears with negative follow-up biopsies: improving cytohistologic correlations. *Diagn*

Cytopathol. 2003;29(4):200-202.

Dr. Crothers, director of cytopathology, Department of Pathology, Walter Reed National Military Medical Center, Bethesda, Md., is chair of the CAP Cytopathology Committee. She was chair of the GCQC2 Cytologic-Histologic Working Group.