The Effect of Continuous Monitoring of Cytologic-Histologic Correlation Data on Cervical Cancer Screening Performance

Stephen S. Raab, MD; Bruce A. Jones, MD; Rhona Souers, BS; Joseph A. Tworek, MD

Objective.—To determine if continuous monitoring of correlation data improves performance.

Design.—Participants in the College of American Pathologists Q-Tracks program (213 laboratories) self-reported the number of Pap test–histologic biopsy correlation discrepancies every quarter for up to 8 years. A mixed linear model determined if the length of participation in the Q-Tracks program was associated with improved performance. Main outcome measures were predictive value of a positive Pap test, Pap test sensitivity, sampling sensitivity, and proportion of positive histologic diagnoses following a Pap test diagnosis of atypical squamous or glandular cells.

Results.—Institutions evaluated 287,570 paired Pap test–histologic correlation specimens and found 98,424 (34.2%) true-positive Pap test correlations, 19,006 (6.6%) false-positive Pap test correlations, and 6,575 (2.3%) false-negative Pap test correlations. The mean predictive value of a positive Pap test, sensitivity, screening and interpretive sensitivity, sampling sensitivity, and proportion of positive histologic diagnoses following a Pap test diagnosis of atypical squamous or glandular cells was 83.6%, 93.7%, 99.2%, 94.2%, 60.3%, and 38.8%, respectively. Longer participation was significantly associated with a higher predictive value of a positive Pap test (P = .01), higher Pap test sensitivity (P = .002), higher Pap test sampling sensitivity (P = .03), and higher proportion of positive histologic diagnoses for a Pap test diagnosis of atypical squamous cells (P < .001).

Conclusions.—Long-term monitoring of cytologic-histologic correlation is associated with improvement in cytologic-histologic correlation performance.

(CArch Pathol Lab Med. 2008;132:16–22)
histologic correlation data resulted in improved laboratory performance over time.

MATERIALS AND METHODS

Study Population
A total of 213 laboratories participated on a year-to-year basis for a variable number of years from January 1, 1999, to December 31, 2006. Participants were grouped into 5 categories according to how long they had participated in this monitor. Twenty-one institutions participated 7 to 8 years, 18 participated 5 to 6 years, 46 participated 3 to 4 years, and 128 participated 1 to 2 years.

Of the participating institutions, 18.8% were government hospitals, 75.6% were nongovernment hospitals, and 5.6% were unknown. Twenty-seven percent of the participating institutions had a bed size between 1 and 250, 49.7% had a bed size between 251 and 500, and 23.3% had a bed size greater than 500.

Definitions
The cytologic-histologic correlation Q-Tracks program10 was based on a Q-Probes study that used standardization definitions and data collection methods. The following definitions of the data elements were provided to all participants:

- **Pap Test Diagnostic Categories.**—Pap tests were categorized using the 2001 Bethesda system classification scheme.
- **Pap Test Error Types.**—**Interpretive Error.**—Cells previously identified and evaluated by the cytopathologist or pathologist, based on the original markers or dots on the slide, are interpreted differently after review.
- **Screening Error.**—Previously unmarked cells are identified during rescreening, resulting in a different interpretation.
- **Adequacy Determination Error.**—On rescreening, a Pap test is unsatisfactory for evaluation.
- **Sampling Error.**—On rescreening, no cytology interpretive, screening, or adequacy errors are evaluated. The review of the Pap test shows that the specimen is adequate and no abnormal cells are present.

**Correlation Categories.**—**False-Negative Pap Test Correlation.**—A negative Pap test with a positive biopsy.

**False-Positive Pap Test Correlation.**—A positive Pap test (atypical Pap test diagnoses excluded) with a negative biopsy. If a positive Pap test was thought to be accurate and the negative histologic specimen in error, laboratories still reported the case as a false-positive Pap correlation, and internally a note was recorded that the histologic diagnosis was incorrect. As the biopsy was used as the gold standard, there was no category for a false-negative histologic correlation.

**True-Positive Pap Test Correlation.**—A positive Pap test (atypical Pap test diagnoses excluded) with a positive biopsy.

Data Collection Process
Data were collected using standardized report forms on a quarterly basis for a maximum of 32 quarters and sent to the CAP. Participants with only a single observation were excluded from this analysis. Participant data were adjusted if there were gaps in their data submission for more than 4 quarters. For each quarter, the laboratory identified patients for whom the primary histologic material was obtained from the uterine cervix, including specimens obtained by biopsy, conization, loop electrosurgical procedure, and cervical curettage (accompanied by a satisfactory unequivocal biopsy specimen). Histologic specimens included those with unequivocal diagnoses. For each included histologic case, the cytopathology files were reviewed for Pap tests (excluding unsatisfactory Pap tests) that had been submitted within 3 months previous to the biopsy or at the time of the biopsy.

A worksheet was provided to each laboratory to track individual Pap test–histologic specimen data. For each case pair, the diagnoses of the histologic specimen and Pap test were compared. If more than 1 Pap test had been performed within 3 months of the biopsy, each Pap test was listed as a separate correlation. If more than 1 biopsy was performed as a result of a Pap test, each biopsy was listed as a separate correlation. The laboratories recorded the number of monthly true-positive, false-positive, and true-negative cases. Laboratories recorded the number of cytology screening, interpretive, both screening and interpretive, adequacy determination, and sampling errors.

Laboratories also tracked the histologic specimen diagnoses for all Pap tests with the diagnoses of atypical squamous cells (ASCs) and atypical glandular cells (AGCs).17 Starting in 2006, laboratories separately tracked the number of high-grade squamous intraepithelial lesion (ASC-H) and ASC—undetermined significance (ASC-US).17

Laboratory demographic and study-specific questionnaires were completed for each year of participation. These questionnaires obtained information regarding institutional characteristics and Pap test–histologic correlation policies and practices.

Analysis
We calculated the following performance metrics:

- **Predictive value of a positive Pap test (%):**
  \[\frac{\text{Number of True-Positive Pap Test Correlations}}{\text{Number of True-Positive and False-Positive Pap Test Correlations}} \times 100\]

- **Sensitivity:**
  \[\frac{\text{Number of True-Positive Pap Test Correlations}}{\text{Number of True-Positive and False-Negative Pap Test Correlations}} \times 100\]

- **Screening and interpretation sensitivity:**
  \[\frac{\text{Number of True-Positive Pap Test Correlations Resulting From Screening and Interpretation Errors}}{\text{Number of True-Positive and False-Negative Pap Test Correlations Resulting From Screening and Interpretation Errors}} \times 100\]

- **Sampling sensitivity:**
  \[\frac{\text{Number of True-Positive Pap Test Correlations Resulting From Screening and Interpretation Errors}}{\text{Number of True-Positive and False-Negative Pap Test Correlations Resulting From Screening and Interpretation Errors}} \times 100\]

- **Proportion follow-up positive histologic specimens for ASC Pap test interpretations:**
  \[\frac{\text{Number of ASC Pap Test Interpretations With a Positive Histologic Diagnosis}}{\text{Number of ASC Pap Test Interpretations With a Corresponding Histologic Specimen}} \times 100\]

- **Proportion follow-up positive histologic specimens for AGC Pap test interpretations:**
  \[\frac{\text{Number of AGC Pap Test Interpretations With a Positive Histologic Diagnosis}}{\text{Number of AGC Pap Test Interpretations With a Corresponding Histologic Specimen}} \times 100\]

The longitudinal data consisted of multiple, equally spaced data points for each institution. These data were referred to as repeated measures. Repeated measures for a single institution are more likely correlated for quarters closer together than for quarters further apart. The variances of repeated measures may change over time. To account for this complicated covariance structure, we fit a mixed linear model.11 The mixed linear model is a generalization of the standard linear model with the generalization being that the data may exhibit correlation and nonconstant variability.

We compared the performance of each institution to prior performance as the number of years of participation increased. The hypothesis that longer participation resulted in improved correlation performance was tested using the mixed linear model PROC MIXED procedure in SAS 9.1 (SAS Inc, Cary, NC).18 An underlying assumption for use of this technique was that the data were normally distributed; as the original proportions were slightly skewed, a natural log transformation was applied to these proportions to obtain an approximately Gaussian distribution.

Arch Pathol Lab Med—Vol 132, January 2008

Cytologic-Histologic Correlation—Raab et al
We tested 7 mixed models using the natural log transformation of the following dependent variables: predictive value of a positive Pap test, sensitivity, screening sensitivity, sampling sensitivity, proportion of false-positive Pap tests, proportion of follow-up positive histologic diagnoses for ASC Pap test interpretations (2006 ASC-US and ASC-H data were pooled to create a single ASC category, similar to the 1999–2005 ASC data), and proportion of follow-up positive histologic diagnoses for AGC Pap test interpretations. Variables analyzed in the mixed linear model are shown in Table 1.

### RESULTS

For the 8 years of the study, institutions evaluated 287,570 paired Pap–histologic correlation specimens and found 98,424 (34.2%) true-positive Pap test correlations, 19,006 (6.6%) false-positive Pap test correlations, and 65,751 (2.3%) false-negative Pap test correlations. The total number of false-negative Pap test screening, interpretive, and adequacy determination errors were 372 (5.7% of all false-negative Pap tests), 280 (4.3%), 102 (1.6%), and 91 (1.4%), respectively. The remaining false-negative Pap tests presumably represent sampling errors.

Table 2 shows the aggregate histologic follow-up diagnoses for women who had a Pap test diagnosis of ASC; 41.9% of the women had a benign histologic diagnosis and 19.7% had a cervical intraepithelial lesion 2 or higher diagnosis. Table 3 shows the aggregate histologic follow-up diagnoses for women who had a Pap test diagnosis of AGC; 64.0% of the women had a benign histologic diagnosis, 22.4% had a cervical intraepithelial lesion 2 or higher diagnosis, and 8.9% had an adenocarcinoma or adenosquamous carcinoma in situ diagnosis.

The mean predictive value of a positive Pap test, sensitivity, screening and interpretive sensitivity, sampling sensitivity, and proportion positive for ASC Pap test interpretations (from years 1999–2005), for ASC-US Pap test interpretations (from 2006), for ASC-H Pap test interpretations (from 2006), and for AGC Pap test interpretations were 83.6%, 93.7%, 99.2%, 94.2%, 60.3%, 62.3%, 71.7%, and 38.8%, respectively. Figure 1 shows the yearly mean predictive value of a positive Pap test, sensitivity, screening and interpretive sensitivity, and sampling sensitivity. Figure 2 shows the yearly mean proportion of women who have follow-up positive histologic specimens for Pap test interpretations of atypical squamous cells and atypical glandular cells.

Longer participation was significantly associated \((P = .01)\) with a higher predictive value of a positive cytology (Figure 3). The institutions that participated for 7 to 8 years showed more improvement in their predictive value of a positive cytology than institutions that participated fewer years. Institution type was a significant factor \((P = .05)\) with predictive values of a positive cytology. Compared with governmental institutions \((n = 48)\), nongovernmental institutions \((n = 152)\) had a higher predictive

### Table 1. Variables Analyzed in the Mixed Linear Model

<table>
<thead>
<tr>
<th>Variable Description</th>
<th>Yes or No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of participating quarters</td>
<td></td>
</tr>
<tr>
<td>Quarter of participation</td>
<td></td>
</tr>
<tr>
<td>Institution is inspected by the College of American Pathologists</td>
<td>yes or no</td>
</tr>
<tr>
<td>Institution is inspected by the Joint Commission</td>
<td>yes or no</td>
</tr>
<tr>
<td>Institution is inspected by the American Association of Blood Banks</td>
<td>yes or no</td>
</tr>
<tr>
<td>Institution is inspected by the Food and Drug Administration</td>
<td>yes or no</td>
</tr>
<tr>
<td>Institution is inspected by the state</td>
<td>yes or no</td>
</tr>
<tr>
<td>Teaching hospital</td>
<td>yes or no</td>
</tr>
<tr>
<td>Nongovernmental institution</td>
<td>yes or no</td>
</tr>
<tr>
<td>Governmental, nonfederal institution</td>
<td>yes or no</td>
</tr>
<tr>
<td>Governmental, federal institution</td>
<td>yes or no</td>
</tr>
<tr>
<td>Institution trains pathology residents</td>
<td>yes or no</td>
</tr>
<tr>
<td>Number of licensed beds</td>
<td></td>
</tr>
</tbody>
</table>

### Table 2. Aggregate Follow-up Histologic Diagnoses for Women Who Have a Papanicolaou Test Diagnosis of Atypical Squamous Cells

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Benign</td>
<td>24,347 (42.8)</td>
<td>2588 (37.1)</td>
<td>396 (29.8)</td>
</tr>
<tr>
<td>Cervical intraepithelial lesion 1</td>
<td>21,404 (37.6)</td>
<td>3210 (46.0)</td>
<td>399 (30.0)</td>
</tr>
<tr>
<td>Cervical intraepithelial lesion 2 or 3</td>
<td>9,513 (16.7)</td>
<td>975 (14.0)</td>
<td>489 (36.8)</td>
</tr>
<tr>
<td>Cervical intraepithelial lesion of indeterminate grade</td>
<td>13,77 (2.4)</td>
<td>144 (2.1)</td>
<td>25 (1.9)</td>
</tr>
<tr>
<td>Adenocarcinoma in situ</td>
<td>44 (0.1)</td>
<td>9 (0.1)</td>
<td>3 (0.2)</td>
</tr>
<tr>
<td>Adenocarcinoma</td>
<td>48 (0.1)</td>
<td>11 (0.2)</td>
<td>6 (0.5)</td>
</tr>
<tr>
<td>Squamous cell carcinoma</td>
<td>87 (0.2)</td>
<td>3 (0.0)</td>
<td>7 (0.5)</td>
</tr>
<tr>
<td>Carcinoma, NOS*</td>
<td>18 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Other malignant diagnosis</td>
<td>53 (0.1)</td>
<td>31 (0.4)</td>
<td>4 (0.3)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>56,891</strong></td>
<td><strong>6971</strong></td>
<td><strong>1329</strong></td>
</tr>
</tbody>
</table>

* NOS indicates not otherwise specified.
value of a positive Pap test (83.1% and 81.5%, respectively).

Longer participation was significantly associated ($P = .002$) with higher Pap test sensitivity (Figure 4). The number of licensed beds was significantly associated ($P < .001$) with Pap test sensitivity. Larger institutions tended to have a lower Pap test sensitivity than smaller institutions; the median Pap test sensitivity of hospitals with bed sizes of 1 to 250, 251 to 500, and more than 500 was 100%, 97.7%, and 96.0%, respectively. Longer participation was significantly associated ($P = .03$) with higher Pap test sampling sensitivity (Figure 5). Longer participation was significantly associated ($P = .04$) with a decrease in the proportion of false-positive Pap tests (Figure 6).

Longer participation in the program also was significantly associated ($P < .001$) with a higher proportion of women who have follow-up positive histologic diagnoses for a Pap test interpretation of ASC (Figure 7). The number of licensed beds was significantly associated ($P = .003$) with the proportion of follow-up positive histologic diagnoses for women who have a Pap test diagnosis of AGC. Compared with smaller institutions, larger institutions tended to have a higher proportion of follow-up positive histologic diagnoses for women who have a Pap test diagnosis of AGC than smaller institutions; the median percent positive AGC diagnoses of hospitals with bed sizes of 1 to 250, 251 to 500, and more than 500 was 25.0%, 38.9%, and 40.0%, respectively.

**COMMENT**

Institutions that participated in the cytologic-histologic correlation Q-Tracks program showed improvement in Pap test performance metrics, indicating that monitoring of correlation data may be used for continuous quality improvement. Based on data that link less than optimal Pap test performance to negative patient outcomes,$2,3$ this study shows that participation in this CAP Q-Tracks monitor may correlate with improved patient care and a lowering of overall health care costs.

Improvement consisted of an increase in Pap test sensitivity, exhibited by fewer false-negative Pap test–histologic specimen correlations. Overall improvement appeared to be driven by improvement in the preanalytic Pap test sampling component, as longer institutional participation also was associated with improved sampling sensitivity. Previous Pap test–histologic correlation studies
have shown that inadequate sampling is the primary root cause of false-negative error,1–5 and targeting the sampling phase of Pap testing theoretically could result in the largest degree of improvement.

Zarbo et al10 reported that institutions that participated in the cytologic-histologic correlation Q-Tracks program for the years 1999 and 2000 did not demonstrate improvement. One explanation for the discrepancy between the...
preliminary data reported by Zarbo et al and the data from the current study is that Pap test sampling may have improved secondary to the introduction of liquid-based technology, which had not been widely implemented prior to 2000. Some studies have shown that liquid-based technologies improved preneoplastic lesion detection and lowered false-negative Pap tests. We hypothesize that institutions that participated the longest and showed the highest degree of improvement reported preliquid- and postliquid-based implementation data, whereas the institutions that participated for shorter periods did not report data from both the preimplementation and postimplementation periods.

Thus, improvement in sampling sensitivity may reflect a national trend secondary to technologic implementation and dissemination rather than individual laboratory data monitoring and feedback. The cytologic-histologic correlation Q-Tracks was not designed to track overall Pap test specificity and therefore it is not possible to determine if improved correlation sensitivity was accompanied by a decrease in correlation specificity. However, the proportion of false-positive Pap test correlations decreased with increased program participation, supporting the hypothesis that increased Pap test detection was not accompanied by Pap test over-diagnosis. In this Q-Tracks program, it also is possible that the reduction in Pap test false-negative errors was a result of biased reporting or the attrition of the more poorly performing laboratories.

These Q-Tracks data show that the majority of false-negative Pap test–histopathologic noncorrelating case pairs are attributed to failures in the specimen sampling phase rather than the specimen interpretation phase of testing. Both of these phases consist of a number of unique steps, and the sampling phase consists of specimen procurement and processing steps. The fact that most false-negative diagnoses are attributable to sampling also indicates that the greatest improvement may be gained by addressing this phase. The application of liquid-based technology standardizes some steps in the procurement and the processing phases of testing. Additional efforts to improve sampling may further improve Pap test performance.
Laboratories that participated longer in this Q-Tracks program also showed a higher proportion of follow-up positive histologic diagnoses for a Pap test diagnosis of ASC. In 2001, the American Society of Cervical Colposcopy and Pathology recommended the use of high-risk human papillomavirus (HPV) testing for women who had an ASC-US diagnosis, as high-risk HPV-positive women were more likely to have cervical disease than high-risk HPV-negative women. Thus, similar to the Pap test sampling sensitivity data, these cytologic-histologic correlation Q-Tracks data may reflect a national trend in improvement secondary to high-risk HPV testing and ASC-US triage.

This study represents the largest published series on the histologic follow-up outcomes of women who have ASC (n = 65191) and AGC (n = 5929). For the cytologic-histologic correlation Q-Tracks program, the separation of the ASC category into ASC-US and ASC-H categories only occurred in 2006 but will serve as a basis to monitor the effectiveness of Bethesda 2001 reporting changes and the use of high-risk HPV testing. The AGC data confirm the findings of numerous smaller studies showing the wide spectrum of diagnoses on histologic follow-up.

The Institute of Medicine report argued for the establishment of national error reporting systems as an initial step to improve patient safety. The CAP created such a system through the Q-Tracks program that has shown the benefit of continuous performance monitoring; studies have shown that increased length of monitor participation has been associated with improvement for a number of metrics, such as frozen section–permanent section correlation, wristband patient identification, blood product wastage, and laboratory specimen acceptability. Limitations in this study include self-reporting bias and the inability to measure a number of clinical and pathology practice variables that may affect data collection and institutional performance. For example, our study was not designed to control for institutional selection bias with the more poorly performing laboratories preferentially dropping out at a far greater rate than the laboratories with better performance. Our study also was not designed for biased reporting of same visit Pap test and histopathologic tissue biopsy diagnoses and for biased reporting for low-grade lesions. A possible explanation for the higher proportion of follow-up positive histologic diagnoses for ASC cases was increased reliance on HPV testing. The correlation process is not nationally standardized, and this cytologic-histologic correlation Q-Tracks program focused on the Pap test and not on the colposcopic examination and biopsy component of cervical cancer screening. Other studies have shown that deficiencies in the colposcopy and biopsy components of testing also will influence the correlation performance data.

In summary, our study of the cytologic-histologic correlation Q-Tracks program has shown a remarkable trend in improved institutional detection of preneoplastic lesions. As the cervical cancer screening field continues to undergo marked transformation, our findings argue for continued monitoring of data to determine how changes in practice are altering cervical cancer screening sensitivity and specificity. Additional studies are needed to determine more fully the causes of these changes.

The College of American Pathologists sponsored this study.

References