Liquid-Based Papanicolaou Tests in Endometrial Carcinoma Diagnosis

Performance, Error Root Cause Analysis, and Quality Improvement

Sharon B. Sams, MD, Heather S. Currens, SCT(ASCP), and Stephen S. Raab, MD

Key Words: Papanicolaou test; Cytopathology; Endometrial adenocarcinoma; Medical error; Quality improvement

DOI: 10.1309/AJCPLFBK1A2XJDQI

Abstract

Recent reports show that the sensitivity of endometrial carcinoma detection on liquid-based Papanicolaou (Pap) tests (88%) is considerably higher than that reported on conventional Pap smears (20%-30%), although few laboratories have corroborated these results. We performed a 5-year retrospective review of all liquid-based Pap tests (n = 69) in women who later were given a diagnosis of endometrial carcinoma, performed error root cause analysis, and developed quality improvement initiatives as a means of error reduction. The original and rescreened Pap test sensitivity rates for endometrial carcinoma were 31.9% and 59.3%, respectively. Root cause analysis showed that poor specimen quality and cognitive failures contributed to a false-negative error in 67% (18/27) and 59% (16/27), respectively, of all cases. System analysis showed that latent factors contributing to error included lack of redundant and educational systems. We conclude that system redesign of liquid-based Pap test screening processes has the potential to improve sensitivity in endometrial carcinoma diagnosis.

Historically, the conventional Papanicolaou (Pap) smear was not considered an effective means to screen for endometrial adenocarcinoma because of low prevalence and low sensitivity.1-3 For example, Mitchell et al3 found that the sensitivity of cervical cytology by smear technique performed within 2 years of the diagnosis of endometrial carcinoma was 28% and concluded that cervical cytology screening would have no major impact on reducing the morbidity or mortality from endometrial carcinoma. A review article by Schnatz et al2 found the prevalence of the diagnosis of atypical glandular cells (AGC) in 24 studies (2,389,206 Pap tests) to be very low (0.29%), and, of these, only 5.2% had malignant follow-up that included cases of endometrial adenocarcinoma.

It is estimated that among women who have endometrial adenocarcinoma, malignant cells are shed in only one third to one half of cases, and it has been inferred that the low sensitivity of detection is primarily secondary to sampling error.4-7 However, few of the early studies that examined the sensitivity of endometrial adenocarcinoma systematically used rigorous root cause analytic techniques to identify the active and latent causes of error. Most studies examining the causes of Pap test false-negative diagnoses of glandular neoplasia focused on endocervical disease.8-10 Lee et al10 retrospectively reviewed cervical smears from 34 women who had a false-negative diagnosis of adenocarcinoma in situ and reported that interpretive errors were a significant factor in the failure of detection. This finding may indicate the presence of a latent problem related to failure of diagnostic criteria development or use for glandular abnormalities.
Studies have shown that liquid-based Pap test technology increased the detection of endometrial adenocarcinoma.\textsuperscript{11-14} Two studies specific to ThinPrep (Hologic, Bedford, MA) technology reported endometrial adenocarcinoma detection sensitivity rates of 65.2\% and 88.3\% compared with 38.6\% with conventional smears.\textsuperscript{12,13} The authors proposed that liquid-based Pap tests improved specimen adequacy and diagnostic yield by removing obscuring blood and inflammation, although rigorous assessments of specimen quality metrics were not performed.

In our study, we set out to confirm the results of these recent findings regarding the effectiveness of liquid-based cytology on endometrial adenocarcinoma detection and extended this work by performing formal root cause analysis to identify the latent and active causes of errors, focusing on the interplay between failures in sampling and diagnostic interpretation. We used our root cause findings to develop quality improvement initiatives as a means to decrease errors in the diagnosis of endometrial adenocarcinoma.

Materials and Methods

Institutional review board approval was obtained for this study (COMIRB protocol, 10-0500).

Case Retrieval

By using our laboratory information system (The Gold Standard, Cortex, Seattle, WA), we performed a 5-year retrospective review of all preceding Pap test reports and slides from women diagnosed with endometrial carcinoma at our institutional anatomic pathology laboratory. We identified a total of 42 women and 69 Pap tests. We retrospectively reviewed 49 Pap tests (71\% of Pap tests were available for review) from 42 women who had a diagnosis of endometrial carcinoma at our institution. All Pap tests reviewed were prepared using liquid-based technology; 36 (73\%) were ThinPrep, and 13 (27\%) were SurePath (BD Diagnostics–TriPath, Burlington, NC). The mean age of the women at the time of diagnosis was 61.2 years (range, 36-91 years).

Assessment of Pap Test Sensitivity for the Detection of Endometrial Adenocarcinoma

Based on the original 2001 Bethesda System interpretation, the Pap tests were stratified into 2 diagnostic categories: negative and positive.\textsuperscript{15} The negative category was composed of all benign results and included the diagnoses of reactive and no evidence of intraepithelial lesion or malignancy. The positive category was composed of all diagnoses that triggered further clinical workup and included all squamous cell abnormalities (ie, atypical squamous cells, low-grade squamous intraepithelial lesion, high-grade squamous intraepithelial lesion) and all glandular cell abnormalities (ie, presence of endometrial cells in a postmenopausal woman, AGC, and adenocarcinoma).

The Pap tests originally classified into the negative category were rescreened by a senior cytotechnologist (H.S.C.) and independently interpreted by a cytopathology fellow (S.B.S.) and an experienced, board-certified cytopathologist (S.S.R.). Based on a consensus of the 2 cytopathologists, the second-review diagnoses were classified into 1 of 3 categories: benign, AGC, or adenocarcinoma. The diagnoses of AGC and adenocarcinoma were considered positive. We calculated the sensitivity of a positive screening result for the original and the second-review diagnoses. We calculated the sensitivity for detection for low-grade tumors (International Federation of Gynecology and Obstetrics [FIGO] I and high-grade tumors, FIGO II or FIGO III) based on the original and second-review diagnoses.\textsuperscript{16-18} In the high-grade group, we separately classified type 2 (serous and clear cell) carcinomas and type 1 (endometrioid) carcinomas.

Root Cause Analysis

We used root cause analysis to determine source of false-negative error in cases with a benign diagnosis on Pap test and endometrial adenocarcinoma diagnosed on follow-up. We performed root cause analysis using 2 methods: (1) No-Blame Box method of the continuous assessment of 2 specimen variables: amount of tumor and specimen quality\textsuperscript{19-21}; and (2) Modified Eindhoven Classification Model for the Medical Event Reporting System for Transfusion Medicine (ECM) involving the assessment of latent and active errors.\textsuperscript{21-24}

In the cytologic-histologic correlation process, cytopathologists generally review slides to assign the error as sampling or interpretation (or both). Our 2-part root cause analysis evaluation focused on the overall process, and we wanted to determine multiple sources of error rather than simply classify error as a clinical procurement and/or an interpretation problem.

No-Blame Box Method

In the No-Blame Box method, a cytopathologist reviews the Pap test slide and classifies the amount of tumor and specimen quality using a pictorial box (Figure 1). The box is divided into 4 quadrants with the amount of tumor depicted on the vertical axis and the degree of quality of the specimen depicted on the horizontal axis. The quality of the specimen is composed of a number of elements, including the overall cellularity, preservation of cells, and presence of obscuring elements (eg, blood, inflammation).\textsuperscript{19-21,25} On review, the cytopathologist records a mark in No-Blame Box that corresponds to her or his assessment of these 2 elements.

This method of root cause analysis provides a measure of interpretability of each specimen. We arbitrarily divided...
the No-Blame Box into 4 quadrants to provide an overall assessment that overlaps with the traditional dichotomous assessment scheme. The 4 quadrants reflect general categories of quality and interpretative failure, recognizing that a more detailed evaluation yields greater granularity of determining error source. For example, specimens classified in quadrant D are considered interpretable for tumor and specimens classified in quadrant B are of high quality and do not contain tumor. In the traditional dichotomous root cause assessment, specimens in quadrants D and B are forms of interpretation and sampling error, respectively. In the No-Blame Box method, the dichotomy is expanded to include quadrants A and C, which consist of poor-quality specimens, without and with the presence of tumor, respectively. The failure to produce a sufficient quality specimen, a specimen with a sufficient amount of tumor, or a specimen in which tumor is interpreted correctly is secondary to factors that affect technical and cognitive skills.

The 2 cytopathologists jointly classified each of the 49 Pap tests in the No-Blame Box, and the number of cases categorized in each quadrant was summed.

The ECM

The ECM classifies error into active causes and latent conditions that may lead to active error.26–28 The ECM classifies error into 3 domains: technical (equipment, forms, and software), organizational (procedures, policies, and protocols), and human (knowledge-based, rule-based, and skill-based) Table 1.22–24 The 3 domains are useful in identifying contributing factors and organizing causes of error and allow for error investigation to focus on system factors rather than entirely on human factors.

By examining the quadrant frequency of No-Blame Box characterizations of error, the 2 cytopathologists used the ECM to reach consensus on the causes of latent conditions and active causes of error. The 2 cytopathologists coded the errors using the ECM and created a table displaying factors that contributed to error. We realize that in much of clinical medicine, the most effective method of performing root cause analysis is immediately after the error occurred.28 The ECM was somewhat limited because the root cause analysis was performed following a lengthy time after the error occurred. However, a benefit of studying overall Pap test performance data was that system issues were better evaluated. The analysis of a population of failures provides greater information on system issues than examining single failure occurrences.

Quality Improvement

Based on the No-Blame Box and ECM root causes of error, we developed quality improvement initiatives that could be used to target specific steps in the laboratory processes that most likely contributed to error. Development was based on Lean methods of identification of work processes of activities, connections, and pathways.29 We recognized that the causes of error were multifactorial, necessitating initiatives that would target different components of work along the entire testing pathway. As preanalytic causes contributed to error, a goal in quality improvement was in identifying factors in the analytic steps representing preanalytic failures. This identification would allow laboratory personnel to handle these specimens differently to mitigate the potential for error.

Results

Of the 69 Pap tests, 27 (39%) preceded a diagnosis of low-grade endometrial carcinoma (FIGO I) and 42 (61%) preceded a diagnosis of high-grade endometrial carcinoma (FIGO II or FIGO III). The sensitivity rates for the detection of endometrial carcinoma based on the original diagnosis were 31.9% overall and 18.5% and 40.4% for low- and high-grade carcinoma, respectively.

Of the 49 Pap tests with an original diagnosis of benign, 27 were available for secondary review. A benign diagnosis was confirmed in 11 of the 27 secondarily reviewed Pap tests, and 16 Pap tests were reassigned into a positive diagnostic category. The sensitivity rates for the detection of endometrial carcinoma following secondary screening and interpretation were 59.3% overall and 57.1% and 61.5% for low- and high-grade carcinoma, respectively.

Table 1 shows the No-Blame Box classification of the 27 reviewed false-negative Pap tests of endometrial carcinoma. In 67% of reviewed Pap tests (18/27), the specimen was of poor quality (quadrants A and C); in only 19% (quadrant D) of reviewed Pap tests was a significant amount of tumor present in a high-quality specimen (5/27). The No-Blame Box classification of the false-negative Pap tests of endometrial...
In the high-grade group, 3 (23%) of 13 Pap tests reviewed were from 1 woman subsequently given a diagnosis of a type 2 malignancy (mixed müllerian tumor with endometrioid, serous, clear cell, and papillary features). The remaining 10 (77%) of 13 Pap tests were from women subsequently given a diagnosis of endometrioid adenocarcinoma (type 1). For the type 2 malignant cases, the No-Blame Box quadrants for Pap test error arose from limitations in specimen quality (quadrant C [once] and quadrant A [twice]).

Table 3 shows the ECM root causes of error for cases classified in each of the 4 No-Blame Box quadrants. Root causes of the laboratory components of error involved in processing, screening, and interpretation phases are listed; preanalytic, clinical components of error were not evaluated in this study.

In the high-grade group, 3 (23%) of 13 Pap tests reviewed were from 1 woman subsequently given a diagnosis of a type 2 malignancy (mixed müllerian tumor with endometrioid, serous, clear cell, and papillary features). The remaining 10 (77%) of 13 Pap tests were from women subsequently given a diagnosis of endometrioid adenocarcinoma (type 1). For the type 2 malignant cases, the No-Blame Box quadrants for Pap test error arose from limitations in specimen quality (quadrant C [once] and quadrant A [twice]).

Table 3 shows the ECM root causes of error for cases classified in each of the 4 No-Blame Box quadrants. Root causes of the laboratory components of error involved in processing, screening, and interpretation phases are listed; preanalytic, clinical components of error were not evaluated in this study.

Table 1
Classification of ECM Root Causes

<table>
<thead>
<tr>
<th>Code</th>
<th>Category</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latent errors</td>
<td>Technical: physical items such as equipment, physical installations, software, materials, labels, and forms</td>
<td>Errors that result from underlying system failures</td>
</tr>
<tr>
<td>TEX</td>
<td>External</td>
<td>Failures beyond the control of the investigating organization</td>
</tr>
<tr>
<td>TO</td>
<td>Design</td>
<td>Inadequate design of equipment, software, or materials; can apply to the design of workspace software packages, forms, and label design</td>
</tr>
<tr>
<td>TC</td>
<td>Construction</td>
<td>Designs that were not constructed properly; examples include incorrect setup and installation of equipment in an inaccessible area</td>
</tr>
<tr>
<td>TM</td>
<td>Materials</td>
<td>Material defects found; examples could be the weld seams on blood bags, defects in label adhesive, or ink smears on preprinted labels or forms</td>
</tr>
<tr>
<td>Organizational</td>
<td>OEX</td>
<td>External</td>
</tr>
<tr>
<td>OP</td>
<td>Protocols/procedures</td>
<td>Quality and availability of protocols that are too complicated, inaccurate, unrealistic, absent, or poorly presented</td>
</tr>
<tr>
<td>OK</td>
<td>Knowledge</td>
<td>Failures resulting from inadequate measures taken to ensure that situational or site-specific knowledge or information is transferred to all new or inexperienced staff</td>
</tr>
<tr>
<td>OM</td>
<td>Management priorities</td>
<td>Internal management decisions in which safety is relegated to an inferior position when there are conflicting demands or objectives; this is a conflict between production needs and safety</td>
</tr>
<tr>
<td>OC</td>
<td>Culture</td>
<td>A collective approach, and its attendant modes, to safety and risk rather than the behavior of just one person; groups might establish their own modes of function as opposed to following prescribed methods</td>
</tr>
<tr>
<td>Active errors</td>
<td>HEX</td>
<td>External</td>
</tr>
<tr>
<td>Knowledge-based behaviors</td>
<td>HKE</td>
<td>External</td>
</tr>
<tr>
<td>Rule-based behaviors</td>
<td>HRQ</td>
<td>Qualifications</td>
</tr>
<tr>
<td>HRC</td>
<td>Coordination</td>
<td>A lack of task coordination within a health care team in an organization</td>
</tr>
<tr>
<td>HRV</td>
<td>Verification</td>
<td>The incorrect or incomplete assessment of a situation, including related conditions of the patient/donor and materials to be used before beginning the task</td>
</tr>
<tr>
<td>HRI</td>
<td>Intervention</td>
<td>Failures that result from faulty task planning and execution; this would be selecting the wrong rule or protocol (planning) or executing the protocol incorrectly (execution)</td>
</tr>
<tr>
<td>HRM</td>
<td>Monitoring</td>
<td>Failures that result from monitoring of process or patient status</td>
</tr>
<tr>
<td>Skill-based behaviors</td>
<td>HSS</td>
<td>Slip</td>
</tr>
<tr>
<td>HST</td>
<td>Tripping</td>
<td>Failures in whole-body movement; these errors are often referred to as “slipping, tripping, or falling”</td>
</tr>
<tr>
<td>Other factors</td>
<td>PRF</td>
<td>Patient-related factors</td>
</tr>
<tr>
<td>Unclassifiable</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ECM, Modified Eindhoven Classification Model for the Medical Event Reporting System for Transfusion Medicine.

Table 2
False-Negative Pap Tests With Follow-up of Adenocarcinoma Categorized by Root Cause Analysis in Each No-Blame Box Quadrant

<table>
<thead>
<tr>
<th>Endometrial Carcinoma Type</th>
<th>No. (%) of Pap Tests/Quadrant</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
</tr>
<tr>
<td>All (n = 27)</td>
<td>7 (26)</td>
</tr>
<tr>
<td>FIGO I (n = 14)</td>
<td>2 (14)</td>
</tr>
<tr>
<td>FIGO II/III (n = 13)</td>
<td>5 (38)</td>
</tr>
</tbody>
</table>

FIGO, International Federation of Gynecology and Obstetrics; Pap, Papanicolaou.
Table 4 shows quality improvement initiatives that were developed based on the ECM classifications of error. These initiatives were developed based on the No-Blame Box quadrant (eg, poor-quality specimen) and encompassed steps in laboratory processing, screening, and final diagnostic interpretation.

Discussion

Our data corroborate the findings of other authors supporting the hypothesis that liquid-based Pap technology may potentially increase the sensitivity of endometrial carcinoma detection.11-13 However, at least in our laboratory, this level of detection was reached only by retrospective review, further signifying that system problems limit initial screening and interpretation processes.

In the United States, 40,083 women per year are given a diagnosis of endometrial carcinoma,30 and a current limitation in disease detection is the lack of a standardized screening method.31,32 Although some authors reported that Pap tests were useful in detecting endometrial carcinoma (eg, vaginal...
The current prevailing notion is that low sensitivity precludes Pap testing as a useful endometrial screening modality.\textsuperscript{1-3} If the data by Zhou et al,\textsuperscript{12} Patel et al,\textsuperscript{11} and our laboratory can be further substantiated, Pap testing could be used as a primary screening tool for endometrial carcinoma, reversing the trend of decreasing Pap testing in perimenopausal and postmenopausal women, as high-risk DNA human papillomavirus testing replaces Pap testing as the primary screening modality.

Zhou et al\textsuperscript{12} and Patel et al\textsuperscript{11} suggested that liquid-based Pap technology increased the sensitivity in detecting endometrial carcinoma by producing a cleaner, higher quality sample. If this were the only factor, we would have expected similar numbers without performing secondary slide review. By using root cause analysis, we identified additional causes of cognitive failures and latent conditions contributing to the original lower sensitivity.

Based on the No-Blame Box method of root cause analysis, the majority of original false-negative FIGO II/III endometrial cancer specimens were of poor quality, whereas a larger proportion of false-negative FIGO I cancer specimens were of good quality. This finding supports the theory that sampling and patient-related issues, such as tumor shedding, have a role in the failure to detect FIGO grade I tumors. In the rescreened population, the sensitivity of detection was similar for low- and high-grade tumors. Therefore, we could expect that if specimen quality issues were adequately addressed, we would see an even greater improvement in sensitivity among patients with FIGO II/III cancer compared with patients with FIGO I cancer. We believe that the poor specimen quality limited the interpretability of Pap tests of high-grade adenocarcinoma, as clusters of malignant cells were difficult to observe or were few. Otherwise, these malignancies presumably would have been diagnosed. For low-grade carcinomas, tumor cells, regardless of volume, were difficult to interpret as the cytologic features were similar to those found in reactive conditions.

The findings in this study highlight the fact that poor specimen quality contributes to difficulties in screening and diagnostic interpretation. Although liquid-based cytology may improve cellular preservation and decrease obscuring factors, some liquid-based Pap tests still are of less-than-optimal quality partly owing to low cellularity, which is more prevalent in the older population. A factor identified in some Pap tests that contributed to poor quality was the presence of lubricant, which, depending on type, compromises the technology, resulting in decreased cellularity. In other Pap tests, clumps of inflammatory cells obscured epithelial cells.

Despite publication of descriptions of specific cytologic criteria for endometrial carcinoma, the usefulness of these criteria in actual practice has not been rigorously evaluated.\textsuperscript{6} In fact, the 2001 Bethesda System does not specify specific criteria for different grades or types of endometrial carcinoma as it does for cervical and endocervical lesions.\textsuperscript{35} Thus, a major challenge in diagnosing endometrial cancers relates to the cognitive tasks involving cancer criteria recognition in poor-quality specimens or in well-differentiated cancers. As many laboratories do not correlate the histologic diagnosis of endometrial adenocarcinoma with preceding Pap tests, an additional limitation in our current system is the failure to recognize that a false-negative Pap test diagnosis occurred, regardless of root cause. A flawed medicolegal system and national leaders who argue against the improbability of detection based on decades-old sensitivity data without formal root cause analysis further precludes greater understanding of the failures or design of quality improvement initiatives.

Our study did not involve a blinded review of the Pap tests, as we focused on determining the root cause of error and not on blinded interpretation to determine sensitivity. A follow-up study will involve blinded review following quality improvement initiatives based on these data to determine the sensitivity of Pap tests in the detection of endometrial adenocarcinoma.

Our proposed quality improvement initiatives are based on system redesign rather than identification and correction of individual failings. As the improvement in Pap test screening is a team effort, clinical and laboratory efforts must jointly target failures in producing a quality specimen. Clinician investigators have reported efforts to improve Pap test quality, using a variety of methods such as Lean and evaluating a number of process steps. The initiatives listed in Table 4 involve laboratory steps that contribute to error. Laboratory personnel must address the cognitive task of identifying poor-quality specimens, which may require a cultural shift in assuming greater responsibility in reporting less-than-optimal specimens. This shift will require revised educational and communication efforts that laboratories would need to undertake. For cognitive failures, we proposed methods of education, redundancy, and improved communication and teamwork tools. For example, laboratory personnel have different levels of expertise in the diagnosis of glandular lesions, and using these experts for local education, rescreening, or prescreening could improve overall laboratory quality in regard to the Pap test diagnosis of endometrial carcinoma.

Root cause analysis of failures in the Pap test diagnosis of endometrial carcinoma indicates a number of active and latent system process problems. Targeting these problems through quality improvement initiatives theoretically could result in improved sensitivity in detection and more widespread use of Pap tests in populations of women at risk for endometrial carcinoma.

From the Department of Pathology, University of Colorado Denver, Aurora, CO.
Address reprint requests to Dr Raab: Dept of Pathology, University of Colorado, 12605 E 16th Ave, Anschutz Inpatient Pavilion, Aurora, CO 80045.

References


