Error Detection in Anatomic Pathology

Richard J. Zarbo, MD, DMD; Frederick A. Meier, MDCM; Stephen S. Raab, MD

- Objectives.—To define the magnitude of error occurring in anatomic pathology, to propose a scheme to classify such errors so their influence on clinical outcomes can be evaluated, and to identify quality assurance procedures able to reduce the frequency of errors.

Design.—(a) Peer-reviewed literature search via PubMed for studies from single institutions and multi-institutional College of American Pathologists Q-Probes studies of anatomic pathology error detection and prevention practices; (b) structured evaluation of defects in surgical pathology reports uncovered in the Department of Pathology and Laboratory Medicine of the Henry Ford Health System in 2001–2003, using a newly validated error taxonomy scheme; and (c) comparative review of anatomic pathology quality assurance procedures proposed to reduce error.

Results.—Marked differences in both definitions of error and pathology practice make comparison of error detection and prevention procedures among publications from individual institutions impossible. Q-Probes studies further suggest that observer redundancy reduces diagnostic variability and interpretive error, which ranges from 1.2 to 50 errors per 1000 cases; however, it is unclear which forms of such redundancy are the most efficient in uncovering diagnostic error. The proposed error taxonomy tested has shown a very good interobserver agreement of 91.4% (κ = 0.8780; 95% confidence limit, 0.8416–0.9144), when applied to amended reports, and suggests a distribution of errors among identification, specimen, interpretation, and reporting variables.

Conclusions.—Presently, there are no standardized tools for defining error in anatomic pathology, so it cannot be reliably measured nor can its clinical impact be assessed. The authors propose a standardized error classification that would permit measurement of error frequencies, clinical impact of errors, and the effect of error reduction and prevention efforts. In particular, the value of double-reading, case conferences, and consultations (the traditional triad of error control in anatomic pathology) awaits objective assessment.

(Arch Pathol Lab Med. 2005;129:1237–1245)

Because of its complex nature, anatomic pathology is prone to error at many steps throughout the testing process. In the complex series of production events on the way to a surgical or cytopathology report, there are few instances of the “autonomation,” mechanized error detection, and safety-inducing “forcing functions” common to industrial production. Professional and technical human interactions are the usual source of quality control and error detection.

A first prerequisite to reducing the incidence of error in anatomic pathology is a reasonably complete and generally consistent set of definitions of the types of problems encountered, so that their magnitude can be gauged. Until now, no such taxonomy has existed. Second, standardized measurement tools, using these definitions, need to not only measure the rates of errors and provide a standardized assessment of their clinical impact, but also test the effects of error reduction and prevention efforts. Our review of the literature discovered no comparable assessment tools; the absence of these tools makes objective assessment of strategies to prevent surgical and cytopathology error impossible. Some strategies have been used to prevent errors in surgical pathology reports. We list these strategies that have been drawn from 2 sources, a relatively small body of published studies on error from individual institutions and the comparative information gleaned from College of American Pathologists (CAP) Q-Probes studies, but have found no reliable assessment of their worth. Two of us (R.J.Z. and F.A.M.) took a step toward providing the first of the 2 missing prerequisites by applying an error taxonomy to our own practice experience in the Department of Pathology and Laboratory Medicine in the Henry Ford Health System (Detroit, Mich).

The 3 of us are now attempting to validate this system in a consortium of institutions studying pathology error and patient safety through a grant funded by the Agency for Healthcare Research and Quality, with the intent of developing the second prerequisite, standardized error monitors. With such monitors in hand, the clinical impact of errors could then be assessed.

DESIGN

Anatomic Pathology Error Detection Studies

Peer-reviewed literature was searched via PubMed for reports from single institutions to examine the mechanisms of error discovery. In each publication, we looked at criteria for defining error and assessing its “severity,”
whether the study was prospective or retrospective, the number of cases reviewed to detect the errors, and the composition of the denominator of cases used to define the error rates.

From the CAP Q-Probes studies of multiple institutions, we examined 4 measures of error: the types and rates of errors detected in amended surgical pathology reports were compared with the types of internal review used in participating departments, and the error types and rates detected by second-pathologist review after case sign-out were compared from usual laboratory quality assurance (QA) and surveillance sources.

Error Taxonomy and Validation
We used a recently developed and validated tool, a taxonomy of anatomic pathology error, to evaluate in a standardized way amended report defects in surgical pathology diagnostic information uncovered in the Department of Pathology and Laboratory Medicine of the Henry Ford Health System in 2001–2003. The approach to error investigation and documentation using that tool is now described (Table 1).

We define 4 general types of errors (Figure), with 3 subtypes in the category of defective interpretation. (1) The first subtype is a false-negative diagnosis or undercall of the extent or severity of a lesion. (2) The second is a false-positive diagnosis or an overcall. (3) The third subtype is misclassification. For example, there is neither an undercall nor an overcall when the pathologist incorrectly labels an entity in the proper category of disease (eg, fibrosarcoma

<table>
<thead>
<tr>
<th>Table 1. Approach to Error Investigation and Documentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Type of error (see Figure)</td>
</tr>
<tr>
<td>- Timing of discovery</td>
</tr>
<tr>
<td>- Discoverer</td>
</tr>
<tr>
<td>- Report revision</td>
</tr>
<tr>
<td>- Mechanism of discovery</td>
</tr>
<tr>
<td>- Outcome of error: initial vs late</td>
</tr>
</tbody>
</table>

Defective Identification
- Patient
- Tissue
- Laterality (right vs. left)
- Anatomic location

Defective Specimen
- Lost specimen, inadequate volume, size, gross description, erroneous measurement or extraneous tissue
- Inadequate representativeness /sampling (tissue, blocks, levels)
- Pertinent ancillary diagnostic study not initially done

Defective Report
- Erroneous/missing non-diagnostic information
- Dictation/Typing error
- Report delivery
- Computer/format, transmission, upload error

Defective Interpretation
- False negative - Undercall
- False positive - Overcall
- Mis-classification: not altering primary or secondary diagnostic characteristics
  - Primary = positive/negative or benign/malignant
  - Secondary = grade, stage, margin, etc.

Error types and test-cycle phases.
rather than malignant fibrous histiocytoma); the alternative designation alters neither the diagnostic primary classification (eg, malignancy) nor secondary diagnostic features (eg, high grade, negative margin) among the characteristics summarized in the report.

The second major category of error is that of defective identification of patient, tissue, or laterality. Such misidentification can take place at any step in the diagnostic process, but typically involves the preanalytic phase of testing. This can involve misidentification of the patient, the origin of the tissue sample itself (eg, stomach vs colon), the anatomic location (eg, ascending vs sigmoid colon), or the laterality of the tissue (eg, right vs left breast).

The third major category of error consists of specimen defects. These include, first of all, lost specimens, but also specimens of inadequate volume or size, as the specimens are submitted in the preanalytic phase, and inadequate gross description or misidentification, extraneous tissue, or inadequate sampling occur in the analytic phase. Analytic-phase specimen defects also include specimens whose representativeness is inadequate or less than optimal at the tissue, block, or slide level, because of an action or inaction taken or not taken in the surgical pathology and histology laboratories. Failure to perform pertinent ancillary studies that would have initially revealed a correct diagnosis is also classified among the subtypes of analytic specimen defects.

The fourth major category of error is that of a defective report. This includes reports with erroneous or missing nondiagnostic information (eg, clinician name, date of procedure), dictation or typing errors, and report format or upload errors. The latter defects arise from the use (or misuse) of computer systems. Defective report errors typically occur in the postanalytic phase of anatomic pathology testing, although absent or incorrect information may have arisen in the preanalytic phase without having been detected or addressed by anyone in pathology until (or after) the report’s preparation and publication.

This scheme incorporates documentation of the type of change made, usually in an amended report, to revise the error after it is discovered. These amendment options include changes in (1) the primary diagnostic characteristics (eg, change from negative to positive, benign to malignant, or inadequate to adequate); (2) the secondary diagnostic characteristics (eg, tumor grade, stage, margin, or node status); (3) diagnostic reclassification (eg, the fibrosarcoma changed to malignant fibrous histiocytoma in which the primary or secondary diagnostic change does not alter the prognostic impact of the classification); (4) patient or specimen reidentification; (5) report of additional specimen sampling that had resulted in the changed report; and (6) other edits of the reports that do not change primary or secondary diagnostic information, patient or specimen identification, or involve specimen characteristics.

Timing of discovery segregates into those cases detected before sign-out (before the case is finalized) and those detected after sign-out (after a report has been produced). For changes detected before sign-out, we define 4 mechanisms of discovery: the effects of (1) additional information or material; (2) intradepartmental review before sign-out or double-read of the current case; (3) preparation for or presentation at a conference or at review with clinician; and (4) an external consultation.

For the revisions after sign-out, we list 5 mechanisms: (1) the responsible pathologist’s review of a recent case without additional information or material; (2) the responsible pathologist’s review of a recent case with additional information or material but without clinician prompting; (3) at preparation or presentation at conference with clinicians (eg, tumor board); (4) clinician-initiated review or reconsideration of a case; and (5) as the result of an external consultation.

The third and last part of the error classification attempts to standardize assessment of outcomes related to anatomic pathology error. Again, temporal considerations force a subdivision of the evaluation into an initial assessment at the time of error discovery and a follow-up as-
assessments 6 months after discovery. The taxonomy of outcome types divide the consequences of results into (a) no impact on care, (b) an impact on care with minimal harm (no morbidity), (c) minor harm (minor morbidity), (d) moderate harm (moderate morbidity), or (e) severe harm (major morbidity or death) (Table 2). Minor morbidity is defined as effects and events that can be demonstrated objectively (fever, thrombocytopenia, wound erythema, swelling, etc), but which do not require hospitalization or surgical intervention. Moderate morbidity includes effects and events that require hospitalization or surgical intervention but not major morbidity, defined as loss of an organ or the function of an organ system (e.g., arm/limb, eye/sight, ear/hearing, speech, or the uterus of a woman of reproductive age) or loss of life.

Finally, from the publications reviewed, an inventory of putative error prevention strategies was accumulated.

RESULTS

1. Literature Review (Table 3)

<table>
<thead>
<tr>
<th>Discrepancy Error Rate</th>
<th>No. of Cases Reviewed</th>
<th>Review Method</th>
<th>Source, y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single Institutions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.26%</td>
<td>5397 SP</td>
<td>Prospective slide double-read</td>
<td>Safrin and Bark,14 1993</td>
</tr>
<tr>
<td>0.2%</td>
<td>Unknown</td>
<td>Questionnaire, 202 pathologists; self-reported errors during 1-y period</td>
<td>Furness and Lauder,15 1997</td>
</tr>
<tr>
<td>0.1%</td>
<td>5000 SP</td>
<td>Retrospective blinded review</td>
<td>Renshaw et al,17 2003</td>
</tr>
<tr>
<td>9% changed diagnosis</td>
<td>. . .</td>
<td>Conference case review</td>
<td>McBroom and Ramsay,6 1993</td>
</tr>
<tr>
<td>10% refined diagnosis</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Multiple Institutions  |                       |               |           |
| 6.7% aggregate, 5.1%   | 6186 SP and CY         | Q-Probes, prospective data collection of up to 100 cases in each laboratory with second pathologist review; 74 laboratories, various QA sources | Raab et al,16 2005 |

* SP indicates surgical pathology; CY, cytopathology; and QA, quality assurance.

2. Double Reader: Specialized Skill Sets.—At Henry Ford Hospital, we have documented the effectiveness of 100% prospective review of all breast cases by a panel of pathologists who have some additional expertise in breast pathology.3 The baseline error rate in terms of amended reports in breast pathology from January 2002 to July 2003 was derived under circumstances of routine practice (ie, voluntary inradepartmental consultation at pathologists’ individual discretion, presentation of malignancies at weekly breast tumor board, and clinician-initiated reviews). During this period, 78,000 surgical pathology cases were seen in the department and 37 amended reports for revised diagnoses were issued. Of these revised diagnoses, 5 (13.5%) were breast cases. Three were false-negative biopsies and 2 were false-positive diagnoses. From August 2003 through June 2004, the process of breast pathology sign-out was changed to daily 100% prospective review by the panel. During this period, 36,000 surgical pathology cases were accessioned, and 18 amended reports were issued for revised diagnoses, but none were breast related. Although no breast diagnoses were changed, the retrospective tumor board review process detected other errors resulting in 4 amended reports regarding errors in secondary diagnostic characteristics related to the assessment of stage (2 cases), margins (1 case), and side of involvement (1 case).

3. Correlation Review.—Although cervical cytohistologic correlations are mandated for American laboratories by the Clinical Laboratory Improvement Amendments of 1988,4 the effect on interpretive error reduction is unproven after 6 years of continuous tracking by the participants in the CAP Q-Tracks program, in which no improvement trends have been documented.5

4. Conference Review.—In 1993, McBroom and Ramsay6 collected data on 416 cases reviewed in 8 conferences during a 14-week period and found that 19% of the diagnoses, or 190 per 1000 cases, were changed after histologic review. Amended diagnoses were split almost equally be-
between altered diagnoses (9%) and refined diagnoses (10%). Eighty-eight percent of the changes were attributed to specialist expertise, and 4.8% were caused by additional information provided by clinicians. Patient management was assessed as unaffected in 92.5% of the 416 conference reviews, 3.8% (n = 16) resulted in major management changes, and 2.9% (n = 12) in minor management changes.

5. Institutional Review.—Institutional consultation of pathology material from patients newly referred to practitioners whose practices the reviewing pathologists support is another time-honored strategy for detecting past errors and preventing future ones. It offers a double-read of cases that have been diagnosed elsewhere before clinical second opinion is rendered or therapy is initiated.

The Hershey Medical Center experience in reviewing case material from all organ systems is that this QA activity produced a 9.1% diagnostic discrepancy rate. These diagnostic differences changed therapy or evaluation in 5.8% of reviewed cases.7 In the published Johns Hopkins Hospital experience of more than 6000 cases, 1.4% of diagnoses changed in ways that caused major modifications of therapy or prognosis.3 However, in both the Hershey and Hopkins series, long-term patient follow-up (ie, outcomes) supported the original pathologist’s diagnosis in roughly 8% of cases, rather than the institutional review pathologist’s second opinion.

Series of organ-specific institutional reviews by specialist pathologists have shown similar rates of changed major diagnoses following institutional review for head and neck (7%),3 prostate (1.3%),8 gynecologic tract (2%),11 and neuropathology (8.8%).12

6. Single-Institution Anatomic Pathology Error Detection: Summary.—Because of variation in definitions and detection methods, the range of error rates in anatomic pathology reported from single institutions is wide, from 1 to 90 per 1000 surgical pathology cases published in the peer-reviewed pathology literature (see Table 3). Variation in definitions of errors and detection techniques makes comparing performance among institutions impossible, so one cannot draw exportable conclusions about successful error reduction strategies. In addition to classifications of error disagreeing from one study to another, calculated rates vary as to whether diagnoses were “changed” or “refined” and according to method of detection, that is, whether cases were derived from pathologists’ self-reporting, prospective double-reading, retrospective blinded review, or case conference review (Table 3).6,13–17

B. Multiple-Institution Anatomic Pathology Error Detection.—The first multi-institutional study of amended anatomic pathology reports carried out in 1996 by the CAP Q-Probes quality improvement program retrospectively reviewed more than 1.66 million surgical pathology case accessions in 359 laboratories.18 The Q-Probes study demonstrated 0.15% median and 0.19% mean rates of changed or amended reports, or 1.5 and 1.9 amended cases per 1000 surgical pathology cases.19 When stratified by slide review practices, no slide review policy resulted in a higher amended report rate of 1.6 per 1000. These various surveillance methods resulted in error detection rates that were not statistically significantly different. A second 2003 Q-Probes study of anatomic pathology discrepancy rates, reported from 74 laboratories, collected data prospectively, examining 6186 surgical and cytopathology cases that were reviewed by a second pathologist after the sign-out of the pathology report. The median error rate detected in this manner was 5.1%, or 51 per 1000 specimens (Table 3).16

The Q-Probes study of 1996 examined amended reports only. The 2003 Q-Probes error study looked at discrepancy rates turned up by routine QA activities in surgical and cytopathology, including internal reviews, such as cytologic-histologic correlations, random review, frozen section—permanent section correlation, extradepartmental consult, intradepartmental consult, conferences, and at clinician request.16 Up to 100 surgical pathology and cytopathology cases, 85% of which were surgical pathology cases, were reviewed after case sign-out in each institution.16

From this review, 1 (5%) in 20 reports had a defect identified. This translates to 50000 defects per million or performance between the 1 and 2 sigma levels. The vast majority of errors detected at post—sign-out review were interpretations changed within the same category of disease (48%), followed by categorical (benign-malignant) interpretation discrepancies (21%), a change in patient or specimen information (9%), and finally by revision of margin status (3.7%).

From the 1996 Q-Probes study, which focused on amended reports, the overall amended report rate was roughly 1.5 per 1000, or roughly 1500 errors per million.13 This rate is significantly a hundred times lower than the 51 per 1000 rate found in the 2003 study. In the 1996 study, the majority of errors involved emendation of significant information that might affect patient management and prognosis “other than the diagnosis.” The second most common change amended diagnostic information itself, and the third most common emendation was of patient identification. The emendation rates in this 1996 study appear to be due largely to “passive” discovery; that is, they were called to pathologists’ attention without being looked for.

In contrast, in the 2003 study when pathologists reviewed 100 cases after sign-out, based on a multitude of QA activities, the “discovery” rate was more than 33 times the rate derived from the 1996 Q-Probes study focused on reports that had been amended. The 74 institutions participating in the 2003 study exhibited a wider range of error detection, from no errors discovered to errors discovered in 21% of retrospectively reviewed cases.16 More active, combined glass slide and report review based on multiple QA initiatives reveals a much higher error rate of 50 errors per 1000 cases, or roughly 50000 errors per million. Unlike the previous study, more comprehensive review most often led to changes in primary diagnosis rather than to changes of nondiagnostic significant information. These 2 well-designed studies illustrate the significant impact of the method of error detection on the rates of errors found.

1. Extradepartmental Consultation.—There are few data on the effectiveness of extradepartmental consultation. The Q-Probes experience is that the extradepartmental consultation process confirms referring pathologists’ original diagnoses 70% of the time and provides significant additional information in 16% of cases.18 Both national and local “experts” are most often used to resolve difficult skin, hematolymphoid, breast, and gastrointestinal cases.19

2. Clinical Impact of Anatomic Pathology Error.—Only the
second of the Q-probes studies examined the clinical consequences of the errors detected. Of 415 discrepant specimens detected in the 2003 Q-Probes study, 379 were evaluated for effect on patient management (Table 4). Pathologist participants described 5.5% of these errors as having had a clinically significant impact on patient management. The outcome assessment by the pathologists participating in this study indicated that only about 1 in 20 pathology errors had, in their estimation, “moderate or marked” impact on clinical management, with nearly 95% resulting in “mild” outcome effect, “near misses,” or “no harm.” The validity of this estimate is limited by a number of uncontrolled factors, which include potential pathologist observer bias and lack of a mandated rigorous investigation. In the context of medical outcomes, chart review is an essential element of assessment, which was not carried out in the Q-Probes studies. Interestingly, this 2003 Q-Probes study showed no statistically significant difference in error frequency by organ type. Others have calculated clinical impact using the number of specimens assessed, rather than the number of errors or discrepant specimens detected, as the denominator. This variation in denominators by itself contributes to the wide range of error rates derived from studies of single institutions. The 2003 Q-Probes study would have had a clinically significant error rate of 0.32% (or 3.2/1000) if the total number of cases reviewed were used for the denominator (Table 4). The confidence in these numbers is further eroded by the lack of standard criteria for outcome assessment.

II. Taxonomy of Error

The wide range of error rates derived from studies of diverse design is not a useful standard against which to make comparisons or with which to judge the effect of interventions on error reduction. One of the authors’ main goals in a multi-institutional investigation (funded by the Agency for Healthcare Research and Quality to study pathology errors and opportunities to improve patient safety) is to develop standardized measurement tools for assessing anatomic pathology errors. As part of this study, we developed and are validating for interinstitutional comparison, a consistent, standardized, relatively comprehensive taxonomy of error in surgical pathology. This scheme started from errors that prompted surgical pathology report emendation. It then looked to see when in the diagnostic process the events causing errors occurred and what the errors were; that is, did it involve diagnostic interpretation, patient or specimen identification, specimen attributes, or report production? It addressed the type of error, the timing of the discovery, the discoverer, and the type of report revision; widened the scope of the investigation to deal with the mechanism of discovery; and provided a classifying framework for evaluating outcomes of error. The outcome assessment is designed to be performed initially, at the time of error discovery, and subsequently at 6 months.

The Figure lists the types of anatomic pathology errors. The central diagram of the figure illustrates the interrelationships of error predicated on root cause analysis. Given the focus on amended reports, some errors that arise from nondiagnostic information defects in the preanalytic aspect of the test cycle may be evident, if initially uncorrected, in the postanalytic test phase as a defective report requiring emendation. The categories of discoverer could be the caregiver, the technical support or clerical support personnel, the pathologist, and/or instrument/computer with designed smart logic. Mechanisms of discovery include those examined in the 2003 Q-Probes studies (eg, routine QA practices; preparation and presentation at clinical conferences and tumor boards; and intradepartmental, internal, extradepartmental, and external consultations). Assessing outcomes has further impressed upon us the importance of agreeing on shared indices of severity of outcomes, which are now incorporated into the assessment.

Validation of the Taxonomy

Our initial evaluation of error identification and classification of 430 amended reports derived from 150000 surgical pathology case accessions using the error taxonomy scheme described herein has shown very good interobserver agreement in error classification of 91.4% (κ = 0.8780; 95% confidence limit, 0.8416–0.9144). In our practice setting, there is a fairly consistent amended report frequency of 2.6 to 3.6 per 1000 during a 3-year study period. More than 1 error has been found in 2% to 10% of these amended reports. Tumor board reviews generated 15% to 18% of the amended reports, and most of these were changes in primary diagnoses based on retrospective clinicopathologic review. The tumor board review process was the most efficient discovery mechanism for interpretive errors, detecting roughly half to three quarters of the interpretive errors documented in amended reports. Importantly, the most common error type resulting in
amended reports was that of wrong identification (27%–38%) related to patient, tissue, laterality, or sample location. The second most common error was a defective report (28%–44%) corrected for erroneous or missing non-diagnostic information, dictation or typing error, or computer format mistake. Diagnostic misinterpretations accounted for a more narrow range of 23% to 28% of the amended reports. Defective specimens were the least common cause of error, accounting for 4% to 10% of the amended reports.

### III. List of Error Reduction/Prevention Techniques

#### From the Literature

Table 5 provides a list of error reduction and prevention techniques found in the literature.

#### COMMENT

Several review and consultation practices are suggested by the laboratory accrediting agencies to reduce error and excess variation in anatomic pathology interpretation, but they provide no specific methodologies. Such methods would include reviewing previous cytologic and/or histologic material from a patient when examining current material; reconciling and documenting significant disparities in intraoperative consultation compared to final pathologic diagnosis, including intradepartmental consultations in the patient's final report; and documenting and maintaining records of extradepartmental consultations. There is no standardization offered of when, how frequently, or how double-reading or second opinion consultations should take place. Furthermore, at this time there are no comprehensive comparisons of the effectiveness of various approaches to double-reading and consultations as a means to reduce error in pathologic interpretation.

In trying to define surgical pathology error, most published studies have focused on diagnostic accuracy. However, our taxonomy finds a wider spectrum of clinically significant errors occurring in surgical pathology with many potential underlying and contributory causes. Examination of the data that the taxonomy uncovers, in root cause analyses, may well reveal that cases of misdiagnosis (the wrong diagnosis for the patient in question) is less an indictment of the pathologist's diagnostic acumen than a problem with patient or specimen misidentification. Correcting this sort of error often requires an investigation into tissue identity, at times even resorting to molecular genetic identity testing of embedded tissues. Diagnostic errors may also result from a pathologist making diagnoses on inadequately sampled tissues, either at the gross or microscopic level, when additional material proves diagnostic. Another defective specimen cause of erroneous

### Table 5. Interpretive Error Reduction Strategies: Types of Observer Redundancy

<table>
<thead>
<tr>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Double-read: general sign-out with intradepartmental consult</td>
</tr>
<tr>
<td>Voluntary, individual diagnostic thresholds</td>
</tr>
<tr>
<td>Mandated by organ, diagnosis, or percent of cases</td>
</tr>
<tr>
<td>Blinded or public review, by individual or panel</td>
</tr>
<tr>
<td>Selected slides or entire case</td>
</tr>
<tr>
<td>Correlation review</td>
</tr>
<tr>
<td>Conference/Tumor Board review</td>
</tr>
<tr>
<td>Extradepartmental consult</td>
</tr>
<tr>
<td>Institutional review (outside cases)</td>
</tr>
</tbody>
</table>

diagnosis is that of extraneous tissue unappreciated as foreign to that sample. These sample-related errors may reflect on the pathologist's practical judgment, but not on his or her knowledge, skills, and abilities as a histopathologist. Lastly, a clinical misinterpretation may result from a pathologist finalizing a report in which he or she failed to catch in proofreading a significant typographic error that lacks diagnostic fidelity with the intended communication. Another report defect that can end in misdiagnosis is failure of a revised diagnosis report to upload into a hospital's computerized information system. The underlying event contributing to the eventual erroneous report is the primary defect in the unfortunate chain of events that may well be considered as the primary cause in the context of subsequent error-prevention root cause analyses. Analysis of misdiagnoses reveals the complexity of errors that occur in all test phases of the surgical pathology testing process.

Two types of antieror interventions have gained currency, at least in North American pathology departments, namely, double-reading and synoptic reporting. One or more elements of observer redundancy may be built into surgical pathology practice to standardize pathologic interpretation through routine case review and diagnostic consultation (Table 5). The other means of achieving standardization and reducing variation is the use of structured data entry and reporting elements, such as checklists and synoptic output forms. Examples of the latter, especially for resected specimens of malignancies, are rapidly becoming standards of practice across the United States.

The most common approach to reducing interpretive error is obtaining a second pathologist's opinion by double-reading of the glass slides constituting a pathology case. This form of intradepartmental consultation is not usually a standardized procedure and may be variously applied in a voluntary manner based on individual diagnostic comfort thresholds. When double-reading is mandated, the mandate tends to be carried out in a bewildering variety of patterns. The double-reading is commonly applied by percentage of cases, by specific organ system (eg, breast lesions), or in specific diagnostic categories (eg, all newly diagnosed malignancies). The impact of both first- and second-reader expertise, the manner in which the review is conducted, and no doubt other variables influence the effectiveness of such error prevention strategies. The advantage to prospective review appears to be the timing of the error correction before erroneous information is communicated, as errors caught before reporting have no opportunity to harm patients.

Several candidate variables may influence error detection by double-reading. A second reviewer may or may not be blinded to the first reviewer's diagnosis. The second review may be undertaken by an individual or panel of double-readers. Review may be based on selected slides or the entire complement of slides in a case. Documentation of intradepartmental consultation in the pathologist's final report, although an accreditation requirement, may also vary. Characteristics that trigger double-reviews also have impact on what the reviews will turn up. Reviews that focus on all new diagnoses of malignancy are directed to minimizing false-positive diagnoses but will find no false-negative diagnoses. Reviews directed to other categories of (benign) diagnoses or that cover all cases in a particular genre can pick up false-negative as well as false-positive diagnoses. Timing of review is another im-

Arch Pathol Lab Med—Vol 129, October 2005

*Error Detection in Anatomic Pathology—Zarbo et al 1243*
portant variable; different mixes of error may be found in cases reviewed before sign-out, in contrast to cases reviewed after sign-out. Most of these review procedures that add an element of redundancy have been adopted in pathology practices on the basis of plausibility (the intuitive likelihood that they will reduce error), but without evidence of relative effectiveness or evaluation of efficiency in terms of time expended and cost incurred per error found.

Some pathology practices are of a size that allows most or all cases of tissue from a specific organ to be directed to subspecialists focused on that organ system for primary review and sign-out. Common areas of partial or full specialization include skin, brain, kidney, liver, and transplant pathology. Full specialization is a cultural system whose adoption greatly influences professional staffing, departmental schedules, and resident education. The simplicity of a single reader has advantages of a pathologist likely to have both a close working relationship and enhanced communication with specialist clinicians and a more complete understanding of specific clinical problems and pathologic correlations that material from the organ system may present. Many clinicians view this approach as the gold standard. A single reader with specialized skill sets using standardized terminology and diagnostic criteria has the advantage of the standardization of one, whereby in the words of Elliott Foucar, “the individuality (variation) of many pathologists is reduced to the individuality of one pathologist.” However, this approach usually comes at the price of staff size and therefore increased direct cost compared to a generalist model. There are no data that demonstrate that specialist sign-out reduces error in pathology.

Correlation reviews present an opportunity to assess diagnostic accuracy by comparing diagnoses between different sampling modalities on the same patient, whether the specimens are synchronously or metachronously obtained. Papaincultur-stained smears and cervical biopsies are examples of frequent specimens collected at the same time; fine-needle aspiration cytology and resection specimens are frequent comparable samples, with the fine-needle aspiration and resection done at different times. There are no data indicating that these legally mandated cytologic comparisons reduce diagnostic errors. Correlations may also be made by integrating information from different testing modalities (eg, biopsy and cytogenetics or molecular genetics). Correlations between morphologic diagnoses and newer molecular analyses have opened a new avenue for error detection; however, the effectiveness of this sort of correlative QA procedure has not yet been documented.

Conference and tumor board reviews constitute another time-honored process in which pathologists review information content of reports and/or review entire glass slide interpretations, often correlated with new clinical and radiographic information. Conferences, especially tumor boards, are often pointed out as QA opportunities that allow pathologists to review accuracy of diagnoses and other report information in a clinical dialog. The effectiveness of this surveillance process depends on the type of cases brought to conference, the quality of the clinical information provided, and the pathologist’s expertise, as well as completeness of the conference-related review and whether or not the review is done by a pathologist different from the initial interpreter. Our experience applying the error taxonomy to our surgical pathology practice appears to affirm the efficiency of the tumor board review process to uncover interpretive errors documented in amended reports.

Whereas clinicians look to their particular pathology departmental expert as the gold standard, pathologists often view extradepartmental experts of choice as the ultimate measures of diagnostic accuracy. It is widely recognized that expert opinions may also differ based on the diagnostic criteria and threshold of the chosen expert. For that reason, many pathologists consistently use the same consultant in a particular area. This practice serves not only for diagnosis verification, but also enables pathologists to calibrate their observations based on the consultant’s diagnostic thresholds. Extradenpartmental consultation is limited as an error surveillance tool, given the focus on out-of-the-ordinary cases rather than on a selection of apparently routine cases that might be false-negative diagnoses on review.

Not all review options beyond cytohistologic correlation and conference reviews are available on-site to small pathology practices. The American Society of Clinical Pathologists consensus conference on second opinions in pathology recommended that pathologists select problem-prone cases for second opinion. Extradenpartmental consultation is more commonly used for this purpose in smaller practice settings than in larger groups, including in-house experts. Cases may also be sent out for referee opinions at the request of clinicians or patients. The status of extradenpartmental consultation as a formal medical consultation is reflected in its own reimbursement code to be distinguished from internal QA activities that are not billable.

As can be seen from this literature review and initial validation of an error taxonomy for anatomic pathology, the value of double-reading, case conferences, and consultations, which intuitively form the traditional triad of error control in anatomic pathology, yet await objective assessment.

This work was funded by the Agency for Healthcare Research and Quality (Rockville, Md), grant RO1HS13321.

References

12. Brunner JM, Inouye L, Fuller GN, Langford LA. Diagnostic discrepancies...


