Comparison of Disagreement and Amendment Rates by Tissue Type and Diagnosis

Identifying Cases for Directed Blinded Review

Andrew A. Renshaw, MD, and Edwin W. Gould, MD

Key Words: Surgical pathology; Cytology; Errors; Quality assurance; Blinded review

Abstract

We sought to determine whether a group of cases that was relatively high in disagreements and subsequent amendments could be identified and targeted for blinded review. During a 4-year period, 8,916 surgical pathology and nongynecologic cytology cases were subjected to blinded review; of these, there were 616 disagreements (6.9%), 69 (0.8%) had subsequent amendments issued, and 33 (0.4%) represented false-negative errors of blinded review. Tissues with the highest amendment rates were breast (4.4%), endocrine (4%), gynecologic (1.8%), and cytology (1.3%). Specimen types with highest amendment rates for specimens with more than 20 cases were breast core biopsies (4.0%) and endometrial curettings (2.1%). Diagnoses were divided into negative (64.7%), malignant (21.4%), nondiagnostic (1%), defined precursor lesions (ie, atypical ductal hyperplasia, tubular adenoma) (9.8%), and atypical or "suspicious" (3.1%). Amendment rates were highest for nondiagnostic material (5%) and atypical/suspicious (2.2%). Reviewing only nondiagnostic and atypical cases would have involved reviewing only 4.0% of cases and detected 14% of amendments. Reviewing all breast, gynecologic, nongynecologic cytology, and endocrine material would have involved reviewing 26.9% of cases and detected 88% of amendments. These data can be used to define material for directed blinded review that is relatively high in potential errors.

Reproducibility in surgical pathology is important for patient management. Reproducibility can be affected by the method that is used to measure it. In particular, nonblinded review can result in significant bias. Nonblinded review is a review of material in which the prior diagnosis and or clinical outcome is known to the reviewer. This is best illustrated by its routine use in gynecologic cytology. Reproducibility in surgical pathology is important for patient management. Reproducibility can be affected by the method that is used to measure it. In particular, nonblinded review can result in significant bias. Nonblinded review is a review of material in which the prior diagnosis and or clinical outcome is known to the reviewer. This is best illustrated by its routine use in gynecologic cytology. In this setting, even though large clinical trials have shown a false-negative rate of no less than 20% for routine screening, review of negative cases (ie, the reviewer already knows the diagnosis was negative) has never shown a false-negative rate of more than 1%. As a result, there has been an increased emphasis on using blinded review (in which the reviewer does not know the previous diagnosis or the clinical outcome) not only in gynecologic cytology but also in surgical pathology. Several studies have shown that disagreement rates using blinded review in surgical pathology range from 4.6% to 9%, and clinically significant disagreement rates or amendment rates range from 0.1% to 4%.3-9

Our department has been using blinded review as a method of quality assurance for 4 years. This method is relatively labor-intensive, and it is not possible to review 100% of material in most departments. As a result, we wanted to determine whether there were subsets of cases that were more likely to result in amendments on subsequent review. To do this, we reviewed the results of blinded review in our department and correlated the disagreement and error rates with the types of tissue, specimen types, and diagnoses.

Materials and Methods

The results of blinded review of specimens interpreted from January 1, 2002, through March 1, 2006, at Baptist...
Hospital of Miami, Miami, FL, were reviewed. Cases originally interpreted by the reviewer (A.A.R.) were excluded. Material included surgical pathology and nongynecologic cytology; gynecologic cytology was excluded.

The method of blinded review has been described. In brief, cases were reviewed within 48 hours of being signed out but without knowledge of the original diagnosis or history. If a disagreement was identified that might be of clinical significance, the case was reviewed with the original pathologist. The original pathologist issued an amendment only if he or she thought such an action was warranted. If there was continued disagreement concerning a case, the case was sent to an outside consultant, and the consultant’s diagnosis was considered the “gold standard.” In general, disagreements that warranted an amendment affected whether the patient required therapy, involved the definitive diagnosis of malignancy, or related to the adequacy of the specimen.

The pathology group whose work was studied included 6 pathologists interpreting primarily hospital-based material. Approximately 20% of the material comes from outpatient sources. The cases undergoing blinded review represented approximately 7.5% of the cases reviewed at this hospital during this period. Although individuals within the group have subspecialty training and sign out more cases within their own specialty, during the study period, all pathologists within the group signed out all types of specimens. The cases were chosen for review for 2 main reasons: (1) review of the work of specific pathologists (because they were relatively new to the group, approximately 1/6 of cases) and (2) review of a random sample of cases from different sign-out services (approximately 5/6 of cases). As such, the organ systems represented in the list are representative of the cases reviewed by the department as a whole, with the exception of skin, which represents 20% to 30% of the cases as a whole but constitutes approximately 50% of the cases in the review group.

Diagnoses were divided into 5 groups: negative, malignant, nondiagnostic, defined precursor lesions (ie, atypical ductal hyperplasia, tubular adenoma, and Barrett esophagus), and atypical or “suspicious.” Negative cases had no malignancy, were adequate, and did not contain anything that was atypical or a defined precursor lesion. Nondiagnostic cases were inadequate owing to the lesion not being sampled (eg, no calcifications in a breast core done for calcifications) or did not meet standard defined adequacy criteria (eg, no pulmonary macrophages in a sputum specimen).

Statistical analysis was performed using a 2-tailed \( \chi^2 \) test.

### Results

A total of 8,916 surgical pathology and nongynecologic cytology cases were subjected to blinded review during a 4-year period, and of these, there were 616 (6.9%) disagreements, 69 cases (0.8%) had subsequent amendments issued, and 33 (0.4%) represented false-negative cases of the blinded review.

Cases originally were divided by tissue type Table II. The tissues with the highest amendment rates were breast (4.4%), endocrine (4%), gynecologic (1.8%), and cytology (1.3%). The rate of amendments in these 4 tissue types all were significantly higher than the rate of amendments in all other tissue types (\( P < .001, P < .001, P < .001, \) and \( P = .006 \), respectively; \( \chi^2 \)). Although endocrine and cytology specimens had relatively high amendment rates, the number of cases reviewed for these 2 categories was relatively small. The specific specimen types with the highest amendment rates for specimens with more than 20 cases were breast core biopsies (4.0%) and endometrial curettages (2.1%) Table II. The most common amendments in breast core biopsies involved atypical ductal hyperplasia (5 cases), lobular carcinoma in

### Table II

Disagreement and Amendment Rates by Tissue Type

<table>
<thead>
<tr>
<th>Tissue Type</th>
<th>Cases</th>
<th>Disagreements</th>
<th>Amendments</th>
<th>Screening Errors on Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bone and soft tissue</td>
<td>186 (2.1)</td>
<td>3 (1.6)</td>
<td>1 (0.5)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Breast</td>
<td>689 (7.7)</td>
<td>90 (13.1)</td>
<td>30 (4.4)</td>
<td>4 (0.6)</td>
</tr>
<tr>
<td>Cytology</td>
<td>151 (1.7)</td>
<td>12 (7.9)</td>
<td>2 (1.3)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Endocrine</td>
<td>56 (0.6)</td>
<td>2 (4)</td>
<td>2 (4)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Gastrointestinal</td>
<td>1,143 (12.8)</td>
<td>36 (3.1)</td>
<td>2 (0.2)</td>
<td>7 (0.6)</td>
</tr>
<tr>
<td>Genitourinary</td>
<td>41 (0.5)</td>
<td>0</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Gynecologic</td>
<td>1,499 (16.5)</td>
<td>101 (6.7)</td>
<td>27 (1.8)</td>
<td>1 (0.1)</td>
</tr>
<tr>
<td>Head and neck</td>
<td>168 (1.9)</td>
<td>18 (10.8)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Hematology</td>
<td>151 (1.7)</td>
<td>15 (9.9)</td>
<td>0 (0.0)</td>
<td>2 (1.3)</td>
</tr>
<tr>
<td>Neuropathology</td>
<td>52 (0.6)</td>
<td>6 (12)</td>
<td>0 (0)</td>
<td>2 (4)</td>
</tr>
<tr>
<td>Pulmonary</td>
<td>52 (0.6)</td>
<td>2 (4)</td>
<td>0 (0)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Skin</td>
<td>4,728 (53.0)</td>
<td>350 (7.4)</td>
<td>5 (0.1)</td>
<td>16 (0.3)</td>
</tr>
<tr>
<td>Total</td>
<td>8,916 (100.0)</td>
<td>616 (6.9)</td>
<td>69 (0.8)</td>
<td>33 (0.4)</td>
</tr>
</tbody>
</table>

* Data are given as number (percentage) of cases. Percentages in the “All Cases” column are based on the total of 8,916 cases; percentages in the remaining columns are based on the row total for the tissue type.
situ/atypical lobular hyperplasia (6 cases), and whether a cellular fibroglandular lesion should be excised to rule out a phyllodes tumor (5 cases). The most common amendments in endometrial curettings involved hormone evaluation or dating (7 cases), adequacy (2 cases), and atypia (2 cases).

Cases also were divided by diagnoses. Diagnoses were divided into negative (64.7%), malignant (21.4%), nondiagnostic (1%), defined precursor lesions (ie, atypical ductal hyperplasia, tubular adenoma) (9.8%), and atypical or suspicious (3.1%). The amendment rates were highest for nondiagnostic material (5%) and atypical or suspicious (2.2%) Table 3, which were significantly higher than the rate for negative specimens ($P < .001$ and $P = .03$, respectively; $\chi^2$).

Reviewing only nondiagnostic and atypical cases would have involved reviewing only 4.4% of cases and detected 14% of amendments. Reviewing all breast, gynecologic, nongynecologic cytology, and endocrine material would have involved reviewing 26.9% of cases and detected 88% of amendments.

**Discussion**

The purpose of this study was simple. Although we believe that blinded review is a very effective and relatively unbiased method to detect disagreements and errors in surgical pathology, we fully recognize that it is extremely labor-intensive, and most laboratories, including our own, will not be able to perform blinded review on all cases. As such, it would be a tremendous benefit to identify the types of cases that would most benefit from this type of review. The data herein strongly suggest that specific tissues, specimens, and diagnoses are more likely than others to result in disagreement and amendment using blinded review. By using these data, individual laboratories can design programs using directed blinded review to identify as many disagreements as possible within the resources available to them. Although the best mix of cases certainly will vary from laboratory to laboratory, the data herein provide an excellent place to start.

The data we present are similar to data that have been presented by others in other settings. Other studies using consultation material also have found a high rate of disagreement in gynecologic tract, breast, and endocrine tissues. In addition, other studies have emphasized hematolymphoid material, genitourinary material, and soft tissue. The present study did not find these areas particularly problematic; however, the number of cases reviewed in these areas was relatively small and may not be entirely representative.

The significance of the type of diagnosis also can vary. Other authors have found, as we have, that atypical cases had a relatively high rate of disagreement. In contrast, other authors have found the majority of the errors they have detected using blinded review to involve malignancy and/or well-defined precursors. Such disagreements suggest that the value of using the diagnosis to determine whether a case needs to be reviewed may vary considerably between practice settings.

Although we believe the results we present are representative of many general surgical pathology settings, the method and the results may vary with the specifics of individual practices. For example, the rates of disagreement may vary with a...
different distribution of tissue types, with the experience and training of the individuals within the group, with the level of subspecialization within the group, and with the level of training of the people doing the review. Nevertheless, the results represent a reasonable starting point to define cases that may benefit from targeted review.

We present data correlating disagreements in blinded review with the tissue type, specimen type, and diagnoses. Such data can be used to define material for directed blinded review, which is relatively high in potential errors.

From the Department of Pathology, Baptist Hospital of Miami, Miami, FL.

Address reprint requests to Dr Renshaw: Dept of Pathology, Baptist Hospital of Miami, 8900 N Kendall Dr, Miami, FL 33176.

References