Impact of Qualified (Indeterminate) Diagnoses on the Accuracy of Renal, Thyroid, and Breast Fine-Needle Aspiration Biopsy

Diane Essex-Sorlie, PhD,1 Cherie L. Penning, MD,2,3 and Gregory G. Freund, MD2,3

Key Words: Diagnostic uncertainty; Test performance characteristics; Fine-needle aspiration biopsy

Abstract

A qualified (indeterminate) diagnosis (QD), such as “suggestive of malignancy,” is thought to complicate patient management by heightening clinical uncertainty. We report that QDs increase the overall effectiveness of renal, thyroid, and breast fine-needle aspiration (FNA) biopsy and that the probability that a qualified diagnosis is negative (QDN) can be predicted by the formula QDN = number of QDs × (proportion of false-negative outcomes/disease prevalence expressed as a proportion). Results of renal (n = 24), thyroid (n = 163), and breast (n = 456) FNA biopsies performed from January 1992 through December 1998 were reviewed and correlated with results of tissue biopsies. For each body site, the FNA biopsies were placed into 1 of 2 diagnostic categories: unqualified diagnoses (UQDs) or QDs. Comparison of test performance characteristics for UQDs only and UQDs combined with QDs demonstrated that inclusion of UQDs increased FNA sensitivity and reduced FNA false-negative diagnoses. More important, the probability that a QD was negative could be predicted from test performance characteristics derived from UQDs.

The fine-needle aspiration (FNA) biopsy is a simple, accurate, fast, and economic method for obtaining tissue for morphologic diagnosis and is very effective in the detection of malignancy.1 Furthermore, it is particularly useful for the diagnosis of metastatic and recurrent malignant neoplasms and for cancer staging. For the clinician, an FNA diagnosis that is positive for malignant neoplasm and is in the correct clinical context may be considered definitive.1 Problems arise, however, when the FNA diagnosis is qualified with terms such as “suggestive of,” “probable,” “possibly,” “likely,” or “questionable for.” In these cases, patient management can be made more difficult by the indeterminacy of the pathologic diagnosis.2 From a patient management standpoint, the clinician is likely to view a qualified (indeterminate) diagnosis (QD) as positive for disease necessitating watchful waiting, repeated FNA biopsy, or surgical intervention.3 At present, little is known about how QDs affect FNA biopsy test performance characteristics or whether a simple method exists to estimate the probability that a QD is indicative of the presence or absence of disease.2,4-6 We show that QDs enhance FNA biopsy test effectiveness by increasing sensitivity and reducing false-negative diagnoses. In addition, we demonstrate that the probability of a QD being negative can be predicted from an examination of the test performance characteristics for FNA biopsies developed from unqualified diagnoses (UQDs).

Methods

Rules for Inclusion of Cases

All renal (n = 56), thyroid (n = 781), and breast (n = 1,200) FNA biopsies performed within the Carle Clinic...
Association/Hospital System (Urbana, IL) from January 1992 through December 1998 were reviewed retrospectively and paired with a corresponding surgically removed or core biopsy specimen (tissue diagnosis). Cases were excluded from analysis for the following reasons: (1) the FNA biopsy and corresponding tissue diagnosis were separated in time by more than 5 months; (2) the FNA biopsy and/or corresponding tissue diagnosis was reported as “quantity not sufficient for diagnosis” or “not diagnostic”; (3) the FNA biopsy and/or corresponding tissue diagnosis were not both performed and interpreted within the Carle Clinic Association/Hospital System; or (4) the tissue diagnosis was qualified. A tissue diagnosis was deemed qualified if it contained the terms “suggestive of,” “probable,” “possible,” “likely,” and/or “questionable for.” Finally, in cases in which multiple FNA biopsies were performed on 1 lesion, the FNA biopsy closest in temporal relationship to the tissue diagnosis was correlated with that FNA biopsy.

Categorization of FNA Biopsy Diagnoses

The aforementioned inclusion criteria resulted in 24 renal, 163 thyroid, and 456 breast FNA biopsies that had a corresponding tissue diagnosis. These FNA biopsies were then divided into 2 categories: (1) FNA biopsies with an unqualified FNA diagnosis or (2) FNA biopsies with a qualified FNA diagnosis. As with tissue diagnoses, FNA diagnoses were deemed qualified if they contained the terms “suggestive of,” “probable,” “possible,” “likely,” and/or “questionable for.” This classification system produced 18 UQDs and 6 QDs for renal FNA biopsies, 100 UQDs and 63 QDs for thyroid FNA biopsies, and 286 UQDs and 170 QDs for breast FNA biopsies. The number of pathologists that interpreted the cases was 8, and the percentages of QDs for each pathologist were 0%, 23%, 27%, 28%, 38%, 40%, 42%, and 68%. For each body site, test performance characteristics were computed from a $2 \times 2$ table in which tissue diagnoses and FNA biopsies were cross-classified. The tissue diagnosis was defined as the “gold standard” against which FNA biopsy diagnoses were evaluated. Standard formulas were used to calculate all test characteristics, eg, sensitivity is the conditional probability of a positive FNA biopsy diagnosis given a positive tissue diagnosis.7

Classification of FNA Biopsy Diagnoses for Use in $2 \times 2$ Tables

All renal, thyroid, and breast FNA biopsy diagnoses were classified as positive, negative, or qualified for malignancy. All tissue diagnoses were classified as positive or negative for malignancy.

For the renal body site, FNA and tissue biopsy diagnoses of renal cell carcinoma, oncocytoma, or angiomyolipoma were classified as positive. No other neoplastic diagnoses occurred in the tabulated cases, and all other diagnoses were categorized as negative.

For thyroid tissue diagnoses, papillary carcinoma, medullary carcinoma, Hürthle cell carcinoma, follicular cell carcinoma, and lymphoma were identified as positive for malignancy. Neoplastic conditions designated as negative for malignancy were follicular adenoma and Hürthle cell adenoma. Thyroid FNA biopsies followed a similar classification scheme, but the diagnosis of follicular or Hürthle cell neoplasm was classified as negative for malignancy if the tissue diagnosis was adenoma and positive for malignancy if the tissue diagnosis was carcinoma. For both FNA and tissue biopsy diagnoses, no other neoplastic conditions were noted in the tabulated cases, and all other diagnoses were grouped as negative.

For the breast body site, the diagnoses classified as positive were duct carcinoma, phyllodes tumor, metastatic carcinoma, and papillary carcinoma. Negative neoplastic diagnoses were hamartoma, fibroadenoma, lipoma, papilloma, and granular cell tumor. As with other sites, no other neoplastic diagnoses occurred in the tabulated cases, and all other FNA and tissue diagnoses were defined as negative.

Results

Test Performance Characteristics of Unqualified and Qualified Renal, Thyroid, and Breast FNA Biopsies

Studies examining the diagnostic utility of FNA biopsies analyze QDs separately, excluding them from the overall reporting of test performance characteristics, or assign them to the positive for disease category.2 To determine whether QDs affect FNA biopsy diagnostic usefulness, test performance characteristics of unqualified and qualified renal, thyroid, and breast FNA biopsies were examined and compared. Table 1 summarizes the test performance characteristics of unqualified renal, thyroid, and breast FNA biopsies.

To include QDs in the overall reporting of test performance characteristics, these indeterminate diagnoses were categorized as shown in Table 2. Because of the rules for defining positive and negative FNA biopsy diagnoses, qualified FNA biopsy diagnoses were identified as positive for disease. Consequently, qualified FNA biopsy diagnoses, when compared with tissue diagnoses, were classified as true-positive or false-positive diagnoses. Test performance characteristics for the renal, thyroid, and breast body sites then were recalculated, with the inclusion of qualified FNA biopsy diagnoses classified accordingly as in Table 2.

Comparison of test performance characteristics using unqualified FNA biopsy diagnoses only with those developed
by combining unqualified and qualified FNA biopsy diagnoses shows that including QDs increased disease prevalence, sensitivity, and false-positive diagnoses and decreased specificity, false-negative diagnoses, positive predictive value, likelihood ratio for a positive test, and likelihood ratio for a negative test. The negative predictive value was unchanged by the inclusion of QDs. Taken together, these findings indicate that the inclusion of QDs in calculating test performance characteristics for FNA biopsies increases test effectiveness if the goal of the FNA biopsy is to rule out more confidently a diagnosis of cancer, i.e., increased sensitivity, decreased false-negative diagnosis, and decreased likelihood ratio for a negative test.

The Number of Qualified FNA Diagnoses That Are Negative by Tissue Biopsy Can Be Predicted From the Test Performance Characteristics of UQDs

An indeterminate FNA biopsy diagnosis, especially one that suggests evidence of disease, can make more difficult the clinical decision of what to do next. Examination of qualified FNA biopsy diagnoses (Table 2) demonstrates that for the renal body site, 17% of QDs correlated with a tissue biopsy negative for disease. For thyroid and breast cases, 22% and 38%, respectively, of qualified FNA biopsies were associated with a negative tissue diagnosis. To determine whether the number of negative qualified FNA biopsy diagnoses could be predicted from examination of FNA biopsy test performance characteristics, UQDs were examined. Table 4 demonstrates that for the renal, thyroid, and breast body sites, 1, 14, and 65 qualified FNA biopsy diagnoses, respectively, were paired with a negative tissue biopsy. The predicted numbers of qualified FNA biopsy diagnoses associated with a negative tissue biopsy were 1, 12, and 74 for renal, thyroid, and breast cases, respectively. These predictions were obtained from the following formula:

\[ QDN = nQDs \times (PFNO/DPEP) \]

where QDN is the number of qualified diagnoses classified as negative, nQDs is the total number of qualified diagnoses, PFNO is the proportion of false-negative outcomes, and DPEP is disease prevalence expressed as a proportion. Furthermore, PFNO and DPEP are taken from FNA biopsy test performance characteristics obtained from UQDs only. When considered together, these findings indicate that qualified FNA biopsy diagnoses are most likely to be positive for disease, i.e., 83% for renal, 78% for thyroid, and 62% for breast cases, and that the number of qualified diagnoses

The following table shows the test performance characteristics of unqualified and qualified FNA biopsies:

**Table 2**

Classification of Qualified Fine-Needle Aspiration (FNA) Biopsy/Surgical (Tissue) Biopsy Pairs

| FNA Biopsy Diagnosis | Surgical Biopsy Diagnosis | Classification of Qualified Diagnosis | No. of Cases
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Qualified</td>
<td>Positive</td>
<td>True positive</td>
<td>5 (83)</td>
</tr>
<tr>
<td>Qualified</td>
<td>Negative</td>
<td>False positive</td>
<td>1 (17)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6 (105)</td>
</tr>
</tbody>
</table>

* Percentage of total qualified diagnoses in parentheses for the body site, e.g., for the renal body site: (5/6) × 100% = 83% of the total qualified diagnoses for this site were true-positive diagnoses and 17% were false-positive diagnoses.
correlated with a negative tissue biopsy can be predicted as follows:

\[
QDN = nQDs \times (PFNO/DPEP)
\]

**Discussion**

These data establish that qualification of FNA diagnoses increases test effectiveness when used to determine the presence or absence of cancer at the 3 body sites examined. Comparison of test performance characteristics for UQDs only and UQDs combined with QDs demonstrates that the inclusion of QDs increased disease prevalence, sensitivity, and false-positive diagnoses. In addition, including QDs decreased false-negative diagnoses, specificity, positive predictive values, likelihood ratio for a positive test, and likelihood ratio for a negative test. Negative predictive values were unchanged (Tables 1 and 3). Thus, including QDs in the calculation of test performance characteristics increases the effectiveness of the FNA biopsy as a diagnostic tool to detect serious treatable disease for which a substantial penalty exists for detection failure. In addition, when sensitivity increases, the proportion of false-negative diagnoses decreases, and this reduction in false-negative diagnoses leads to a drop in the likelihood ratio for a negative test. Therefore, when a negative FNA biopsy occurs, the clinician can exclude more confidently a diagnosis of cancer.9,10

Overall, the FNA biopsy percentage of QDs for the body sites examined was 25% for renal, 39% for thyroid, and 37% for breast. In contrast, the QD rates for the surgical diagnoses at these same sites were 0% for renal, 0.4% for thyroid, and 0.92% for breast. Examination of FNA biopsy QDs showed that 83%, 78%, and 62% of qualified renal, thyroid, and breast diagnoses were positive for disease (Table 2).

While the literature is replete with information illustrating the test performance characteristics of the FNA biopsy when diagnostic results are quantified in a binary fashion as positive or negative for disease,1 information on diagnoses that are qualified by the pathologist is lacking. Most studies addressing diagnostic indeterminacy in FNA biopsies address thyroid, but in these studies, indeterminacy equates with “follicular neoplasm,” in which the FNA biopsy cannot determine reliably whether a neoplasm is an adenoma or carcinoma.11-15 In our study, the diagnosis of follicular neoplasm was not considered a QD but rather a true-positive or a true-negative diagnosis depending on its correlation with the surgical specimen.

In breast and renal FNA biopsies, diagnoses deemed atypical or suggestive of malignancy constitute between 8% and 15% of the cases examined,16-20 but these studies do not clearly address qualifying words in the FNA biopsy report and, therefore, underestimate their use.

In our study, all diagnoses with modifying words were considered nondefinitive, even if the qualification favored a benign diagnosis. This rationale is supported because only FNA biopsy cases with matched tissue biopsies were included for study. Thus, regardless of the degree of pathologic certainty, if qualification leads to excision, clinical uncertainty remained.

Because QDs were defined as true-positive or false-positive diagnoses, sensitivity was increased when they were included in calculating test performance characteristics. Specificity, however, was decreased markedly. Importantly, if QDs had been defined differently (see following text), their effect on sensitivity and specificity would be altered because their 2 × 2 table assignments would be changed. Therefore, owing to the potential problem of classifying QDs consistently between different institutions, we conclude that they should be examined separately from non-QDs to provide comparable clinical information.

One of the most commonly reported methods for attempting to assess the nonbinary nature of FNA biopsy specimens is the use of likelihood ratios.2,5,21 Likelihood ratios summarize the same kind of information as sensitivity and specificity and can be used to calculate the probability of disease after a positive or negative test, ie, pretest odds × likelihood ratio = posttest odds. However, it is important to note that pretest probabilities contain the same information as prevalence; likelihood ratios contain the same information as specificity and sensitivity; and posttest probabilities contain the same information as positive predictive value.7,9

We show herein that inclusion of QDs when examining test performance characteristics of the FNA biopsy lowers the likelihood ratio for a positive test and the likelihood ratio for a negative test (Table 3). This was expected...
because QDs were defined as true-positive diagnoses or false-positive diagnoses (Table 2), and the degree to which false-positive diagnoses increased when QDs were included in the calculation of test performance far exceeded that of true-positive diagnoses. These findings underscore the problem of QDs, because to use them in calculations of test performance characteristics, they should be classified in a binary fashion. In addition, no established rules exist for defining QDs in this way. We defined QDs with positive surgical diagnoses as true-positive diagnoses and those with negative diagnoses as false-positive diagnoses. It could be argued that QDs with positive surgical diagnoses are false-negative diagnoses because the diagnosis of cancer was never made and QDs with negative surgical diagnoses are true-negative diagnoses because the diagnosis of malignancy was not stated definitively. Added to this complexity is the use of intermediate indeterminate diagnoses, such as atypical squamous cells of undetermined significance (ASCUS) favor reactive and ASCUS favor dysplasia, used in Papanicolaou testing. Therefore, we conclude that simple disclosure of the probability that a QD is positive for disease or negative as in Table 2 might more accurately communicate to the clinician what outcomes are expected when a diagnosis is qualified.

Finally, we found that the probability of a QD being negative could be predicted from test performance characteristics derived from UQDs using the formula QDN = nQDs × (PFNO/DPEP) (Table 4). This finding was unexpected because it had not been reported previously. Additional study is recommended to assess the generalizability of this outcome, but these results suggest that there is a fixed quantifiable relationship among disease prevalence, false-negative diagnoses, and diagnosis qualification and that the probability that a QD represents malignancy is an intrinsic quality of the test used.

**References**