Inter-Observer Variation in the Pathologic Identification of Extranodal Extension in Nodal Metastasis from Papillary Thyroid Carcinoma


**Background:** Extranodal extension (ENE) in lymph node metastases has been shown to worsen the prognosis of papillary thyroid cancer (PTC). Despite the clinical significance of ENE, there are no stringent criteria for its microscopic diagnosis, and its identification is subject to inter-observer variability. The objective of this study was to determine the level of agreement among expert pathologists in the identification of ENE in PTC cases.

**Methods:** Eleven expert pathologists from the United States, Italy, and Canada were asked to review 61 scanned slides of representative permanent sections of PTC specimens from Mount Sinai Beth Israel Medical Center in New York. Each slide was evaluated for the presence of ENE. The pathologists were also asked to report the criteria they use to identify ENE.

**Results:** The overall strength of agreement in identifying ENE was only fair ($\kappa = 0.35$), and the proportion of observed agreement was 0.68. The proportions of observed agreement for the identification of perinodal structures (fat, nerve, skeletal, and thick-walled vessel involvement) ranged from 0.61 to 0.997.

**Conclusions:** Overall agreement for the identification of ENE is poor. The lack of agreement results from both variation in pathologists’ identification of features and disagreement on the histologic criteria for ENE. This lack of concordance may help explain some of the discordant information regarding prognosis in clinical studies when this feature is identified.

**Introduction**

The overall prognosis for the vast majority of patients with papillary thyroid carcinoma (PTC) is excellent. However, the risks of recurrence and mortality are significantly increased for those with advanced stages of disease. Nodal involvement is one factor that influences the staging of thyroid disease in the American Joint Committee on Cancer (AJCC) system for tumor nodal disease and distant metastasis (TNM) staging. A patient with PTC who is >45 years of age is upstaged from Stage I to III or IV if nodal metastases are identified (patients are upstaged to Stage III if lymph node metastases are present in level 6 of the neck and upstaged to Stage IV if lymph node metastases are found in the lateral compartment [levels 1–5] or the mediastinum [level 7]). In the 2009 guidelines, the American Thyroid Association...
(ATA) proposed a three-tiered stratification system for the risk of recurrence, with the categories of low, intermediate, and high risk of recurrence. Patients with cervical lymph node metastases are placed in the intermediate risk category (1).

It has long been recognized that there is a wide spectrum of nodal characteristics in patients with PTC that ranges from micrometastases (manifesting as single isolated, or “naked,” psammoma bodies, single viable tumor cells, or small clusters of tumor cells) to enlarged, macroscopic nodes demonstrating extensive extranodal extension (ENE) (2). Thus, increasingly, studies suggest that future PTC staging systems ought to account for various nodal characteristics, such as size, number, location, and ENE (3–6). In 2012, Randolph et al. published a meta-analysis that addressed the impact of different nodal characteristics on the risk of disease recurrence (4). In that study, the characteristics considered most significant included (i) whether a node is clinically evident, (ii) the number of positive nodes, (iii) the size of the largest node, and (iv) the presence of ENE (4).

The overall rate of ENE in patients with PTC is reported to be between 22% and 45% (6–8). ENE is a strong and independent predictor of disease recurrence and disease-specific survival (6,8–10), and a predictor of distant metastases (11). Of note, the study by Randolph et al. suggested that ENE is the strongest prognostic factor among all lymph node parameters (even more so than location of nodal involvement) (6).

In a recent study, it was found that 62% of patients undergoing prophylactic central compartment dissections had one or more pathologically positive nodes, and 14% of these patients had lymph nodes that demonstrated ENE (12). However, the exact incidence of ENE in metastatic lymph nodes in PTC cases is difficult to estimate, since performing a prophylactic central neck dissection in the absence of clinically evident nodal disease is controversial.

Even though ENE has become an increasingly recognized prognostic factor in PTC, it must be acknowledged that there are no stringent criteria for its microscopic diagnosis, and thus its identification can be subject to inter-observer variability (3). The most recent College of American Pathologists Protocol for the Examination of Specimens From Patients With carcinomas of the Thyroid Gland (2014) notes the importance of documenting the presence or absence of ENE (13). However, histologic criteria for identifying ENE are not defined (13).

To the authors’ knowledge, there are no reported studies analyzing the degree of concordance among pathologists in the assessment of ENE. This study seeks to evaluate the degree of inter-observer variability in the histopathologic identification of ENE, as well as the current criteria used by pathologists for its identification. As the prognostic significance of ENE is increasingly recognized and incorporated into PTC staging systems, it is important to understand the variability in its histopathologic identification.

### Materials and Methods

This study was performed with approval by the Mount Sinai Beth Israel Medical Center Institutional Review Board. Eleven expert pathologists from the United States, Canada, and Italy were asked to review 61 scanned slides of representative permanent sections of PTC specimens from Mount Sinai Beth Israel Medical Center in New York. Slides were selected that were potentially controversial with respect to the presence of ENE (determined by a senior pathologist, B.M.W.). All slides were prepared with standard hematoxylin and eosin (H&E) stain. The slides were digitally scanned at 40x using the Aperio ScanScope AT Turbo. All 11 pathologists evaluated each case.

Pathologists reviewed each slide and designated the specimen as demonstrating ENE or having no evidence of ENE. In addition, pathologists were asked to evaluate specifically for perinodal involvement of fat, skeletal muscle, nerve, and thick-walled vascular structures. A survey was used to determine the specific criteria used by each pathologist for diagnosing ENE.

The kappa coefficient was computed to assess consistency in the pathologists’ categorization of each slide and the features evaluated. All statistical analyses were done with R statistical package v3.1.1 (R Development Core Team, Vienna, Austria).

### Results

Overall, the proportion of observed agreement for the histologic diagnosis of ENE was 0.68 and produced a kappa coefficient of 0.35, indicating that the strength of agreement was fair (Table 1). The proportion of observed agreement for the identification of perinodal structures was lowest for perinodal fat involvement (0.61), and highest for perinodal nerve and muscle involvement (0.98 and 0.997).

All of the surveyed pathologists reported that they use perinodal fat involvement as a histologic criterion for diagnosing ENE (Table 2). Skeletal muscle, nerve, and thick-walled vessel involvement are also taken into consideration to varying degrees. Seven pathologists reported use of all four

<table>
<thead>
<tr>
<th>Kappa coefficient</th>
<th>Strength of agreement</th>
<th>Proportion of observed agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presence of ENE</td>
<td>0.35</td>
<td>Fair</td>
</tr>
<tr>
<td>Perinodal fat involvement</td>
<td>0.2</td>
<td>Slight/fair</td>
</tr>
<tr>
<td>Perinodal nerve involvement</td>
<td>0.09</td>
<td>Slight</td>
</tr>
<tr>
<td>Perinodal skeletal muscle involvement</td>
<td>0</td>
<td>Slight</td>
</tr>
<tr>
<td>Perinodal thick-walled vessel involvement</td>
<td>0.02</td>
<td>Slight</td>
</tr>
</tbody>
</table>

*Kappa agreement: <0, less than chance agreement; 0.00–0.20, slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; 0.81–0.99, almost perfect agreement.

ENE, extranodal extension.
perinodal structures as part of the diagnostic criteria for ENE, and the remaining four pathologists use between one and three structures. The most common additional criterion used by pathologists to diagnose ENE is “tumor beyond nodal capsule,” which was reported by 8/11 of the pathologists in this study. Other criteria utilized by pathologists include “fibrosis with desmoplasia outside node,” “extrinsic extension into vessels, not tumor emboli,” and “desmoplastic stromal reaction outside node.”

**Discussion**

As indicated by a kappa coefficient of 0.35, the overall strength of inter-pathologist agreement among 11 experts for the histologic diagnosis of ENE in PTC is only fair. The low degree of concordance indicates that the pathologic diagnosis of ENE can be inconsistent and prone to subjectivity.

Raters indicated that perinodal nerve, skeletal muscle, and thick-walled vessel involvement were not observed for the vast majority of slides. While the kappa coefficients are low for the identification of these features, this is due to the low incidences of positive identification. Thus, for this analysis, the proportions of observed agreement for perinodal nerve, skeletal muscle, and thick-walled vessel involvement, which were 0.98, 0.997, and 0.87 respectively, are the more accurate representation of inter-pathologist agreement.

The concordance for identifying individual structure involvement was higher than concordance for the identification of ENE. While the proportion of observed agreement for the identification of ENE was 0.68, the proportion of observed agreement for the identification of perinodal fat, nerve, skeletal muscle, and thick-walled vessel involvement were 0.61, 0.98, 0.997, and 0.87, respectively. These results indicate that a disagreement over the proper criteria for identifying ENE is what leads to the poor inter-pathologist concordance in the overall diagnosis of ENE. These findings suggest that the standardization of the histologic criteria for ENE through establishment of consensus pathologic guidelines should improve uniformity in the diagnosis of ENE.

There are several potential limitations of this study. First, the sample size included only 11 pathologists. Only senior pathologists from various medical centers in the United States, Canada, and Italy considered to be experts in thyroid pathology were invited to partake in the study. This naturally biases the group toward pathologists who are more experienced with ENE and PTC, and who would not necessarily be representative of all pathologists who might review nodal specimens from patients with PTC. In fact, one might surmise that the degree of concordance for less experienced pathologists would be less than that determined in this study. A second study designed to engage pathologists in community hospitals would help to clarify this point. The second limitation is that the pathologists were only reviewing a single imaged slice from each specimen. In practice, pathologists often review multiple sections and/or levels. Since there is no standardized protocol for the extent to which a lymph node should be sectioned, it is likely that the presence of ENE is under-reported by virtue of the fact that more extensive sectioning may identify areas of extranodal spread.

The presence of ENE is a significant prognostic factor in thyroid cancer, is an important component in the assignment of risk of recurrent disease, and may be taken into account in

---

**Table 2. Histologic Criteria Utilized by Pathologists for Diagnosis of ENE**

<table>
<thead>
<tr>
<th>Pathologist 1</th>
<th>Pathologist 2</th>
<th>Pathologist 3</th>
<th>Pathologist 4</th>
<th>Pathologist 5</th>
<th>Pathologist 6</th>
<th>Pathologist 7</th>
<th>Pathologist 8</th>
<th>Pathologist 9</th>
<th>Pathologist 10</th>
<th>Pathologist 11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perinodal fat involvement</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Perinodal skeletal muscle involvement</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Perinodal nerve involvement</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Perinodal thick-walled vessel involvement</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Other: Tumor beyond nodal capsule</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Fibrosis with desmoplasia outside node</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Desmoplastic stromal reaction outside node</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Tumor beyond nodal capsule</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Tumor beyond nodal capsule involving extranodal tissue</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Tumor beyond nodal capsule involving tumor emboli</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>

818 DU ET AL.
the decision to administer radioactive iodine therapy. Thus, clinicians must understand the lack of concordance in pathologic diagnosis of ENE. The lack of concordance in identifying ENE could also greatly impact the interpretation of clinical studies that assess the influence of ENE on disease prognosis.

In conclusion, this study evaluates inter-observer concordance in the histologic identification of ENE, a finding that is increasingly considered important in the staging and treatment of PTC. Overall agreement on the identification of ENE is poor, likely caused by disagreement in the histologic criteria used for the diagnosis of ENE. The lack of concordance in the diagnosis of ENE has potential implications for both disease prognosis and patient care, and it needs to be recognized by clinicians treating PTC patients in order to allow appropriate assessment of studies that include ENE as a prognostic variable. Clinicians should be cognizant of this variability when reviewing the pathology reports on their patients.

Acknowledgments

We would like to acknowledge the generous support of the THANC Foundation and the Mount Sinai Health System.

Author Disclosure Statement

No competing financial interests exist.

References


4. Randolphp GW, Duh Q-Y, Heller KS, LiVolsi VA, Mandel SJ, Steward DL, Tufano RP, Tuttle RM; American Thyroid Association Surgical Affairs Committee’s Taskforce on Thyroid Cancer Nodal Surgery 2012 The prognostic significance of nodal metastases from papillary thyroid carcinoma can be stratified based on the size and number of metastatic lymph nodes, as well as the presence of extra-nodal extension. Thyroid 22:1144–1152.


Address correspondence to:
Meghan E. Rowe, BA
Thyroid, Head and Neck Cancer (THANC) Foundation
10 Union Square East, Suite 5B
New York, NY 10003

E-mail: mrowe@thancfoundation.org