Initial Performance of NI-RADS to Predict Residual or Recurrent Head and Neck Squamous Cell Carcinoma


ABSTRACT

BACKGROUND AND PURPOSE: The Head and Neck Imaging Reporting and Data System (NI-RADS) surveillance template for head and neck cancer includes a numeric assessment of suspicion for recurrence (1–4) for the primary site and neck. Category 1 indicates no evidence of recurrence; category 2, low suspicion of recurrence; category 3, high suspicion of recurrence; and category 4, known recurrence. Our purpose was to evaluate the performance of the NI-RADS scoring system to predict local and regional disease recurrence or persistence.

MATERIALS AND METHODS: This study was classified as a quality-improvement project by the institutional review board. A retrospective database search yielded 500 consecutive cases interpreted using the NI-RADS template. Cases without a numeric score, non-squamous cell carcinoma primary tumors, and primary squamous cell carcinoma outside the head and neck were excluded. The electronic medical record was reviewed to determine the subsequent management, pathology results, and outcome of clinical and radiologic follow-up.

RESULTS: A total of 318 scans and 618 targets (314 primary targets and 304 nodal targets) met the inclusion criteria. Among the 618 targets, 85.4% were scored NI-RADS 1; 9.4% were scored NI-RADS 2; and 5.2% were scored NI-RADS 3. The rates of positive disease were 3.79%, 17.2%, and 59.4% for each NI-RADS category, respectively. Univariate association analysis demonstrated a strong association between the NI-RADS score and ultimate disease recurrence, with $P < .001$ for primary and regional sites.

CONCLUSIONS: The baseline performance of NI-RADS was good, demonstrating significant discrimination among the categories 1–3 for predicting disease.

ABBREVIATIONS: AUC = area under the curve; CECT = contrast-enhanced CT; H&N = head and neck; NI-RADS = Head and Neck Imaging Reporting and Data System; ROC = receiver operating characteristic
Management: Routine surveillance (6 months typically, see “Materials and Methods”)

- **Category 2: Low suspicion of recurrence**
  Imaging: Ill-defined abnormality with only mild enhancement and/or FDG uptake
  Management: Direct inspection for mucosal findings or short (3-month) follow-up with CECT or PET for deep findings

- **Category 3: High suspicion of recurrence**
  Imaging: Discrete, new, or enlarging lesions with intense enhancement and/or focal FDG uptake
  Management: Biopsy

- **Category 4: Known recurrence, pathologically proved or definitive radiologic or clinical progression.**

NI-RADS categories 1–4 are the same for CECT or CECT/PET, but the linked management recommendations and lexicon are slightly different (because FDG avidity is included in the latter). Furthermore, the first version of NI-RADS category 2 contains subcategories that also address lesion size and location (superficial or deep): 2a, superficial/mucosal surface; 2b, deep abnormality of < 1 cm; and 2c, deep abnormality of > 1 cm. These subcategories are useful to direct management, for example superficial mucosal abnormalities are amenable to direct inspection. Size criteria were not added to predict disease but rather to avoid biopsy in this indeterminate category unless immediate management depended on the biopsy. This template-driven approach reflects common language to promote collaboration between radiologists and referring providers, data-driven optimization of H&N cancer imaging, and greater direct engagement with patients.

An obstacle to improvement in value-based performance and direct patient reporting is lack of a data-driven standard surveillance imaging algorithm. PET/CECT at 12 weeks is often the first posttreatment study, though a recent study suggests that it can be performed at 8 weeks. At our institution, patients with advanced H&N cancer are scanned with CECT/PET at 12 weeks as a baseline. If the findings are negative, they undergo CECT alone 6 months later, and if these findings are negative, they undergo CECT alone 12 months later. Although studies have investigated PET/CT for surveillance, ordering practices among treating physicians remain variable. The 2015 National Comprehensive Cancer Network recommendations advocate imaging within 6 months for T3/4 primary tumors or N2/3 nodal disease and then additional imaging only for new signs/symptoms, smoking, or areas inaccessible to clinical inspection (the latter being arbitrary and difficult to apply). Yet, 79% of H&N cancer surgeons self-reported using PET/CT for asymptomatic patients. Given this variation in practice, it is critical to have measurable categories to correlate with outcomes to develop a data-driven universal surveillance algorithm.

NI-RADS allows H&N radiologists to perform structured radiologic-pathologic correlation and to determine accuracy, prognostic value, and interobserver agreement in contrast to subjective interpretations that provide no data for retrospective analysis. Our objective was to determine the initial performance of the NI-RADS scoring system to predict tumor recurrence or persistence in patients treated for squamous cell carcinoma of the H&N undergoing imaging surveillance.

**MATERIALS AND METHODS**

This study was designated a Quality Improvement project by our institutional review board at the Emory University School of Medicine. An electronic medical record search from June 12, 2014 to January 28, 2015 yielded 500 consecutive neck CECT examinations interpreted with the NI-RADS template, including patients with a variety of tumor types and primary locations. The following was gathered from a review of the electronic medical record:

- 1) Age and sex
- 2) Date and site of original diagnosis
- 3) Human papillomavirus status
- 4) Initial tumor stage (Tumor, Node, Metastasis)
- 5) Treatment (ie, surgery, chemotherapy, radiation)
- 6) Date of the index scan
- 7) Type of index scan (CECT alone versus CECT with PET scan)
- 8) First posttreatment scan versus subsequent surveillance scan
- 9) Length of imaging and clinical follow-up

**Inclusion Criteria**

1) Treated primary H&N squamous cell carcinoma.
2) CECT and/or CECT/PET for surveillance.
3) NI-RADS template used for interpretation.
4) Total of 402 scans met the inclusion criteria.

**Exclusion Criteria**

1) Insufficient outcomes data to determine positive or negative disease.
2) Score of “4, known recurrence” because recurrence had already been proved before the scan. However, it is possible to have a score of 4 for the primary site and 1, 2, or 3 for lymph nodes (or vice versa). Thus, an outcome could still be determined for nodes so that each scan had 2 possible sites for target abnormalities (primary and neck).
3) Multiple scans in the same patient if there were back-to-back scores of 1 for both primary and neck. In this case, subsequent index scans were excluded because the final outcome of “recurrence or not” for the primary or neck would be the same for these 2 data points.
4) Either dataset yielded 287 patients, 318 scans, and 618 total targets (314 primary targets and 304 nodal targets) for which outcomes could be determined.

**Surveillance Algorithm and Image Interpretation**

At our institution, all patients with advanced H&N cancer (almost all patients except those with T1 N0 disease) are scanned with CECT/PET at 12-week baseline, and if the findings are negative, they undergo a CECT alone 6 months later. If these findings are negative, they undergo a CECT 12 months later. All NI-RADS surveillance scans were interpreted prospectively using the template by 1 of 4 dedicated H&N neuroradiologists (30, 15, 10, and 9...
FDG uptake was determined. Univariate association between recurrence and scan score (1–3) was measured by \( \chi^2 \) test and the nonparametric Fisher exact test. The same analysis was repeated for primary site, lymph node, and their combination separately. The overall performance of discrimination of the scan score on recurrence status (yes versus no) was measured as the area under curve (AUC) by receiver operating characteristic (ROC) analysis with 95% confidence intervals. The sensitivity and specificity at each cut-point of the scan score were presented accordingly for score 1 versus 2–3 and for scores 1–2 versus 3. Additionally, the same ROC analyses for subgroup performed for CECT alone versus CECT + PET and for the first posttreatment examination versus the subsequent surveillance examination were explored. The interobserver agreement was measured by \( \kappa \) statistics among 40 scans for primary and neck sites by 2 graders. The statistical significance level was set at \( P < .05 \), and analyses were conducted in SAS 9.4 (SAS Institute, Cary, North Carolina).

RESULTS
Of the 318 examinations, there were 221 CECTs alone (69.5%) and 97 CECT/PETs (30.5%). Sixty studies (16.4%) were initial baseline posttreatment examinations (performed at 12 weeks); the remainder were follow-up examinations during routine surveillance per protocol.

Median imaging follow-up after the index scan was 51 weeks; median clinical follow-up was 54 weeks. The distribution of tumor site and initial stage (when known) is outlined in Table 1. Primary tumors of the oropharynx were the largest group (43.2%), followed by tumors of the oral cavity (25.4%) and larynx (22.3%). At the primary site, almost one-third had moderately advanced (T4a) disease (32.8%). More than half had at least N2 nodal disease (54.7%), Distant metastatic disease at initial staging was rare (2.1%).

Table 1: Tumor site and initial stage (patient level)

<table>
<thead>
<tr>
<th>Primary site</th>
<th>% (No.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oropharynx</td>
<td>43.2% (124)</td>
</tr>
<tr>
<td>Larynx</td>
<td>22.3% (64)</td>
</tr>
<tr>
<td>Oral cavity</td>
<td>25.4% (73)</td>
</tr>
<tr>
<td>Hypopharynx</td>
<td>4.2% (12)</td>
</tr>
<tr>
<td>Skin</td>
<td>2.1% (6)</td>
</tr>
<tr>
<td>Unknown</td>
<td>2.8% (8)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nodal stage</th>
<th>% (No.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nx</td>
<td>4.5% (13)</td>
</tr>
<tr>
<td>N0</td>
<td>28.9% (83)</td>
</tr>
<tr>
<td>N1</td>
<td>11.8% (34)</td>
</tr>
<tr>
<td>N2a</td>
<td>3.1% (9)</td>
</tr>
<tr>
<td>N2b</td>
<td>33.8% (97)</td>
</tr>
<tr>
<td>N2c</td>
<td>16.4% (47)</td>
</tr>
<tr>
<td>N3</td>
<td>1.4% (4)</td>
</tr>
<tr>
<td>Distant stage</td>
<td>97.9% (281)</td>
</tr>
<tr>
<td>M0</td>
<td>2.1% (6)</td>
</tr>
<tr>
<td>M1</td>
<td></td>
</tr>
</tbody>
</table>

Interobserver Agreement
The interobserver agreement determined by \( \kappa \) statistics after review of 40 scans (80 targets) by 2 graders was very good, 0.821 (95% CI, 0.657–0.986) with \( P < .001 \).

Incidence of Disease Recurrence/Persistence Based on NI-RADS Score
The incidence of recurrence for each NI-RADS category is detailed in Table 2. Overall, the incidence of tumor persistence/recurrence was 7.9%, with an 8.9% (28/314) recurrence rate at the primary site and a 6.9% (21/304) regional nodal recurrence rate.

Image Acquisition
All PET/CT imaging followed standard protocol, and was performed on GE Discover 600 and 690 PET/CT scanners (GE Healthcare, Milwaukee, Wisconsin). Patients fasted for 6 hours before the scan, and serum glucose concentration was obtained immediately before FDG administration. The examination was deferred if glucose was >200 mg/dL. Combined PET/CT from the skull vertex through the midthigh was obtained 1 hour after intravenous administration of 10–14 mCi of FDG. Helical noncontrast CT from the vertex through midthigh was performed before PET for attenuation correction and anatomic localization. A CECT of the neck with the arms down was performed following PET. Our split-bolus technique used 110 mL of intravenous iopamidol (Isovue-370; Bracco, Princeton, New Jersey), with 55 mL injected first at 2.5 mL/s, a 40-second delay, then another 55 mL at the same rate, with a total scan delay of 90 seconds. We acquired axial images from the frontal sinuses through the mediastinum at the same rate, with a total scan delay of 90 seconds. We acquired axial images from the frontal sinuses through the mediastinum at 1.25-mm section thickness; pitch, 0.984:1; gantry rotation, 0.7 seconds; FOV, 25 cm; 120 kV(peak); and Smart milliampere with 1.25-mm section thickness; pitch, 0.984:1; gantry rotation, 0.7 seconds; FOV, 25 cm; 120 kV(peak); and Smart milliampere.

Statistical Methods
Univariate association between recurrence and scan score (1–3) was determined by a subjective dichotomous analysis of intense FDG uptake was determined.
NI-RADS 1. Five hundred twenty-eight of 618 targets (85.4%) were scored “NI-RADS 1, no evidence of recurrence” with only 3.8% having recurrent disease during the follow-up. When considered separately, the recurrence rate for primary and nodal NI-RADS 1 scores was similar (3.5% and 4.0%, respectively).

NI-RADS 2. Fifty-eight of 618 targets (9.4%) were scored “NI-RADS 2, questionable recurrence” and had a higher overall rate of recurrence of 17.2%, with similar rates for primary and nodes separately (18.4% versus 15.0%). Of 58/618 category 2 lesions, there were 38 primary site category 2 lesions (27/38 “2a,” 7/38 “2b,” and 4/38 “2c”). Seven of 38 underwent biopsy with 5/7 positive, and 2 patients had imaging progression, for a total recurrence of 7/38 (18.4%). There were 20 neck category 2 lesions (15 “2a” and 5 “2b”) with 2/20 having pathology-proven recurrence and 1/20 with clinical disease progression, for a total recurrence rate of 3/20 (15.0%) for the neck. There was no difference in the rate of positive disease based on lesion size within the confines of this small sample size.

NI-RADS 3. Thirty-two of 618 targets (5.2%) were “NI-RADS 3, highly suspicious for recurrence” and had the highest overall recurrence rate of 59.4%, with a 54.6% recurrence at the primary site and a 70.0% rate at nodes. Of the 32/618 category 3 lesions, there were 22 primary site lesions and 10 neck lesions. Twenty-two of 32 category 3 targets had pathologic confirmation of disease presence or absence. The remaining 10/32 did not have pathologic confirmation because it would not affect management (n = 7) or the ultrasound or CT correlate of the suspected lesion could not be found when biopsy was attempted (n = 3). Eight of these 10 (80.0%) had clinical or radiologic evidence for recurrence (7 primary site lesions and 1 nodal site), defined as progression at the target site on imaging or clinically obvious tumor.

NI-RADS Performance

Univariate association analysis demonstrated a strong association between the NI-RADS score and ultimate disease persistence/recurrence, with \( P < .001 \) for primary site, lymph node scores, and combined scores. ROC curves for NI-RADS performance at the primary site (Fig 1), lymph nodes (Fig 2), and combined (Fig 3) were obtained and reflect an overall good performance. For the primary site ROC curve (Fig 1) (AUC = 0.787; 95% CI, 0.691–0.881), \( P < .001 \) indicated a good performance of the NI-RADS score to discriminate primary site recurrence versus no recurrence (an AUC value of 1 indicates a perfect discrimination, and an AUC value of 0.5 indicates no use). For lymph nodes, the AUC of 0.712 and an AUC of 0.756 for combined primary and nodal sites indicated good overall performance of this rating scale.

Subgroup Analysis of CECT Alone versus PET/CECT

A subgroup analysis was undertaken comparing the performance of CECT alone versus CECT + PET/CT (Table 3). The overall
The recurrence rate in these 2 groups was similar (7.0% versus 10.1%). Although there was no statistical difference in overall performance of NI-RADS for CECT (AUC = 0.779) versus CECT/PET (AUC = 0.709), a NI-RADS 3 on CECT alone was more likely to correctly identify recurrence (primary or nodal) compared with a NI-RADS 3 on CECT/PET (91.7% versus 40.0%).

**Table 3: CECT alone versus CECT with PET/CT**

<table>
<thead>
<tr>
<th>NI-RADS</th>
<th>CECT alone</th>
<th>CECT + PET/CT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.1% (12/385)</td>
<td>5.6% (8/143)</td>
</tr>
<tr>
<td>2</td>
<td>21.9% (7/32)</td>
<td>11.5% (3/26)</td>
</tr>
<tr>
<td>3</td>
<td>91.7% (11/12)</td>
<td>40.0% (8/20)</td>
</tr>
<tr>
<td>Combined, all categories</td>
<td>7.0% (30/429)</td>
<td>10.1% (19/189)</td>
</tr>
</tbody>
</table>

**Table 4: Initial posttreatment versus subsequent follow-up**

<table>
<thead>
<tr>
<th>NI-RADS</th>
<th>Initial Posttreatment</th>
<th>Follow-Up</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.7% (5/88)</td>
<td>3.4% (75/440)</td>
</tr>
<tr>
<td>2</td>
<td>20.0% (4/20)</td>
<td>15.8% (6/38)</td>
</tr>
<tr>
<td>3</td>
<td>50.0% (4/8)</td>
<td>62.5% (75/24)</td>
</tr>
<tr>
<td>Combined, all categories</td>
<td>11.2% (13/116)</td>
<td>7.2% (36/502)</td>
</tr>
</tbody>
</table>

**Subgroup Analysis of Initial Posttreatment Study versus Subsequent Studies**

An additional subgroup analysis compared the performance on initial posttreatment studies with performance on subsequent follow-up (Table 4). While there was no statistical difference in the overall performance of NI-RADS in initial posttreatment surveillance (AUC = 0.729) versus subsequent scans (AUC = 0.760), the recurrence rate for NI-RADS 3 in initial posttreatment was greater for the initial baseline scan group (5.7%, 5/88) compared with the subsequent follow-up examination group (3.4%, 15/440). This difference was even more pronounced when looking at the primary site alone (9.3% versus 2.4%). As expected, the incidence of positive disease was also greater in the initial posttreatment group versus surveillance studies (11.2% versus 7.2%).

**DISCUSSION**

The baseline performance of NI-RADS demonstrated significant discrimination between groups, with disease recurrence/persistence rates of 3.8% for NI-RADS 1, 17.2% for NI-RADS 2, and 59.4% for NI-RADS 3. A strong association between score and positive disease was found for primary site, lymph nodes, and all targets combined, and ROC analysis also demonstrated clinically significant and accurate performance in these categories. While adding additional NI-RADS categories may improve ROC performance, the simplicity of the current scale is appropriate for the limited management options: routine surveillance, shorter interval follow-up, additional PET/CT imaging, mucosal inspection, or biopsy.

Because all these patients are part of our institutional surveillance program, with routine follow-up, it is reasonable for the specificity to be high and sensitivity lower. In fact, size cutoffs were set for the “ill-defined” or “questionable” NI-RADS 2 lesions to avoid low-yield, difficult, dangerous, or likely nondiagnostic disease.
biopsies for an intermediate suspicion lesion in a complex post-treatment neck if short-interval follow-up or PET would be a viable option. Because size was not helpful in predicting recurrence or for practical management decisions, our revised NIRS category 2 does not have size criteria and management consists of earlier follow-up or PET. Biopsy recommendations are reserved for category 3 lesions only. In our experience with highly suspicious lesions, size rarely determines recurrence, compared with enhancement characteristics, morphology, interval change, and FDG uptake, which have been incorporated into the NIRS lexicon.

Our NIRS template has been useful in daily clinical practice. For the primary site, the 2a category is used for low-suspicion superficial mucosal lesions with a linked recommendation of direct inspection. Focal asymmetric enhancement and FDG uptake are not an uncommon finding in posttreatment imaging and could represent benign mucositis or early recurrence/persistence. Although many mucosal abnormalities are false-positives (Fig 4), we are able to identify mucosal recurrences, especially in the post-radiated larynx where abnormalities may be subtle on CECT (Fig 5A). In this clinical scenario, the fused PET images (Fig 5B) help direct inspection and biopsy. For the primary site, the 2b category is used for deep, ill-defined, nondiscrete, low-suspicion lesions with only mild FDG uptake (if combined with PET) (Fig 6). In practice, most category 2 lesions are managed with short-term follow-up rather than biopsy because clinicians and patients were comfortable waiting. Short-term follow-up has become our official recommendation in this category because size criteria were removed.

Finally, NIRS 3 is reserved for a discrete, nodular, robustly enhancing lesion (Fig 7A) with marked FDG uptake if PET was also performed (Fig 7B), and the recommendation is for biopsy. In the neck, NIRS 3 is a new or enlarging lymph node (Fig 8A) with marked FDG uptake if PET is combined (Fig 8B). The positive predictive value for NIRS 3 primary site lesions was lower (54.6%) than for the neck (70%); this finding likely reflects the more complex posttreatment imaging appearance at the primary site. Overall, we believe that the NIRS template yielded a reasonable rate of recommending biopsy. Only 32 of 618 possible targets (5.2%) were scored category 3 with biopsy recommendation, balanced against a relatively high positive predictive value (54.6% for the primary site, 70% for the neck).

Our subgroup analyses highlight areas for future study. Although the numbers are small, our data suggest that CECT alone may be more specific because the rate of true persistence/recurrence was much higher for a NIRS 3 for CECT (91.7%) alone versus CECT/PET (40%). We also separated our scans into the first posttreatment baseline at 3 months and subsequent surveillance studies to understand the variation of NIRS performance at different time points. As expected, a NIRS 1 on a subsequent follow-up examination had a higher negative predictive value than on the initial posttreatment examination. This is valu-

FIG 6. NIRS primary site category 2b: ill-defined asymmetric soft tissue. T4N0 oral cavity squamous cell carcinoma. CECT shows asymmetric full soft tissue around fibular reconstruction of the mandible (arrow). The linked management recommendation is shorter interval surveillance. Repeat CECT at 3 months showed no interval change (not shown). Subsequent clinical follow-up also demonstrated improvement and no disease recurrence.

FIG 7. NIRS primary site category 3: discrete enhancing lesion. T4a larynx squamous cell carcinoma, status post total laryngectomy, bilateral neck dissection, and chemoradiotherapy. A, CECT shows a 1-cm discrete rounded hyperenhancing nodule along the lateral border of neopharynx, deep to the flap (arrow). B, Fused PET images show focal high FDG uptake (arrow). This was given a category 3 score, and endoscopic biopsy demonstrated recurrence.

FIG 8. NIRS neck category 3: new or enlarged lymph node. T2N0 oral cavity squamous cell carcinoma status post resection, neck dissection, and adjuvant radiation therapy. A, CECT at 6-month intervals shows enlarging left level 1B lymph node with necrosis (arrows). B, Fused PET images show marked focal FDG uptake (arrow). Revision neck dissection was positive for disease recurrence.
able in providing guidance to patients regarding their risk of disease at different time points.

Finally, NI-RADS provides a meaningful framework for discussion of results with patients. For example, a patient with a NI-RADS 2 on surveillance imaging has a chance of recurrence of roughly 17.2%. We can also reassure patients with NI-RADS 1 that their overall recurrence rate is low (3.8%). There is an opportunity to understand the negative and positive predictive values of NI-RAD scores at different time points in further subpopulation studies. For example, our subgroup analysis and comparison of NI-RADS 1 score primary site recurrence rates for the initial post-treatment examinations versus subsequent examinations found a difference (9.3% versus 2.4%, $P = .047$), but the overall numbers were small because the incidence of recurrence in this group was so low. This analysis suggests that a NI-RADS 1 on the initial baseline posttreatment examination is not as reassuring as a NI-RADS 1 on subsequent surveillance examinations.

CONCLUSIONS
The performance of NI-RADS was good, demonstrating significant discrimination between groups, with positive disease rates of 3.8% for NI-RADS 1, 17.2% for NI-RADS 2, and 59.4% for NI-RADS 3. Standardization of linked management recommendations and correlation with patient outcomes should validate performance and highlight the added value of radiologists in patient care.

REFERENCES