AJNR

Inter- and Intrareader Agreement of NI-RADS in the Interpretation of Surveillance Contrast-Enhanced CT after Treatment of Oral Cavity and Oropharyngeal Squamous Cell Carcinoma

F.H.J. Elsholtz, S.-R. Ro, S. Shnayien, C. Erxleben, H.-C. Bauknecht, J. Lenk, L.-A. Schaafs, B. Hamm and S.M. Niehues

Inter- and Intrareader Agreement of NI-RADS in the Interpretation of Surveillance Contrast-Enhanced CT after Treatment of Oral Cavity and Oropharyngeal Squamous Cell Carcinoma

F.H.J. Elsholtz, S.-R. Ro, S. Shnayien, C. Erxleben, H.-C. Bauknecht, J. Lenk, L.-A. Schaafs, B. Hamm, and S.M. Niehues

ABSTRACT

BACKGROUND AND PURPOSE: The Neck Imaging Reporting and Data System was introduced to assess the probability of recurrence in surveillance imaging after treatment of head and neck cancer. This study investigated inter- and intrareader agreement in interpreting contrast-enhanced CT after treatment of oral cavity and oropharyngeal squamous cell carcinoma.

MATERIALS AND METHODS: This retrospective study analyzed CT datasets of 101 patients. Four radiologists provided the Neck Imaging Reporting and Data System reports for the primary site and neck (cervical lymph nodes). The Kendall's coefficient of concordance (W), Fleiss κ (κ_F), the Kendall's rank correlation coefficient (τ_B), and weighted κ statistics (κ_w) were calculated to assess inter- and intrareader agreement.

RESULTS: Overall, interreader agreement was strong or moderate for both the primary site (W = 0.74, κ_F = 0.48) and the neck (W = 0.80, κ_F = 0.50), depending on the statistics applied. Interreader agreement was higher in patients with proved recurrence at the primary site (W = 0.96 versus 0.56, κ_F = 0.65 versus 0.30) or in the neck (W = 0.78 versus 0.56, κ_F = 0.41 versus 0.29). Intrareader agreement was moderate to strong or almost perfect at the primary site (range τ_B = 0.67–0.82, κ_w = 0.85–0.96) and strong or almost perfect in the neck (range τ_B = 0.76–0.86, κ_w = 0.89–0.95).

CONCLUSIONS: The Neck Imaging Reporting and Data System used for surveillance contrast-enhanced CT after treatment of oral cavity and oropharyngeal squamous cell carcinoma provides acceptable score reproducibility with limitations in patients with posttherapeutic changes but no cancer recurrence.

ABBREVIATIONS: BI = Breast Imaging; CECT = contrast-enhanced CT; LI = Liver Imaging; NI = Neck Imaging; OCSCC = oral cavity squamous cell carcinoma; OPSCC = oropharyngeal squamous cell carcinoma; PI = Prostate Imaging; RADS = Reporting and Data System

Oral cavity squamous cell carcinoma (OCSCC) is the most common malignancy of the head and neck but might soon be overtaken by oropharyngeal squamous cell carcinoma (OPSCC), whose incidence is rapidly rising, mainly because its occurrence is related to the human papillomavirus.[1-3] Smoking and alcohol use are outstanding risk factors with synergistic effects.[4] While some authors use OCSCC for cancers in both locations, we think it is important to separate them. The oral cavity is separated from the oropharynx by the junction of the hard and soft palates above and the

circumvallate papillae located at the transition from the anterior two-thirds to the posterior third of the tongue below.[5]

After completion of curative treatment for OCSCC or OPSCC, patients are enrolled in a program of continuous surveillance imaging and clinical examinations. Surveillance imaging can be performed using CT, MR imaging, or PET/CT and PET/MR imaging.[5,6] Radiologists interpreting posttherapeutic imaging studies in these patients typically focus on the detection of submucosal recurrence at the primary cancer site and the identification of suspicious lymph nodes in the neck. Mucosal recurrence might also be seen in surveillance imaging but is a domain of referring clinicians. Especially in patients who underwent high-dose radiation therapy, the best surveillance can be ensured with a combination of clinical examinations, high-resolution imaging, and possibly endoscopy.[7]

Interpretation of posttherapeutic neck imaging studies in these patients is often challenging for radiologists. In this setting, nonstandardized framing is the common way to rate the probability of cancer recurrence. Reporting and Data Systems (RADS) provide standardized terminology and guidance toward a final
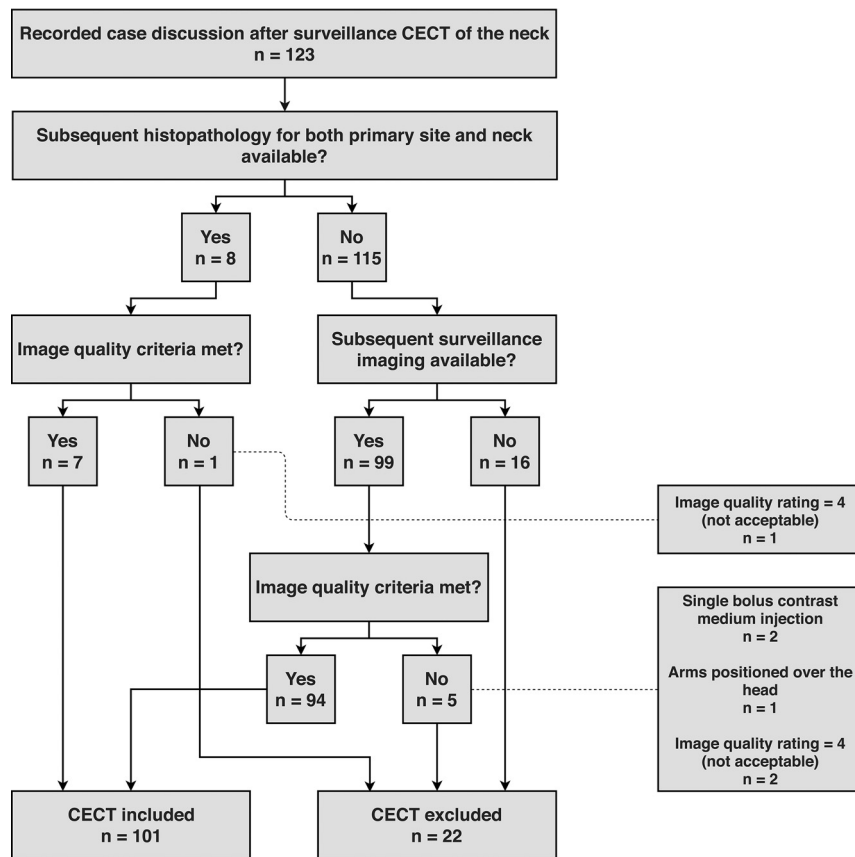
FIG 1. Flow chart of study participants.

score reflecting the probability of malignancy in patients enrolled in cancer surveillance programs. Following the introduction of such a system for breast imaging (BI-RADS) in 1997, several RADS for different organs and body regions (eg, PI-RADS for the prostate and LI-RADS for the liver) have been published and also become highly appreciated by referring clinicians, not in the least because they improve comparability and reproducibility.[8-10]

In 2016, the Neck Imaging Reporting and Data System (NI-RADS) was introduced by the American College of Radiology and has shown a promising initial performance.[11-13] Defined features and findings lead to a numeric value that reflects the probability of cancer recurrence and is directly linked to recommendations for measures to be taken for further patient management.

The major motivation to perform this study was to test NI-RADS for its reliability in interpreting contrast-enhanced CT (CECT), which is, by far, the most common technique used for the surveillance of patients with head and neck cancer in our institution, to obtain evidence to support its implementation as a reporting standard for imaging studies and discussion of findings with referring physicians from the department of oral and maxillofacial surgery.

MATERIALS AND METHODS
Patient Population
This retrospective study was approved by our institutional review board, and written informed consent was obtained from all

patients. In the records of our weekly interdisciplinary conferences (of radiologists and oral and maxillofacial surgeons) held between June 2017 and July 2019, we identified 123 consecutive patients for whom CECT studies performed at our department or by an external institution were available, and 101 patients (41 women, 60 men; median age, 64 years) were finally included in this study. A flow chart of participants is provided in Fig 1. A total of 202 target sites (primary cancer site and neck for each patient) were evaluated. Of the patients included, 72 had OCSCC localized in the mouth floor ($n = 22$), the anterior two-thirds of the tongue ($n = 19$), the hard palate ($n = 3$), and the gingival, labial, or buccal mucosa ($n = 28$). Twenty-nine patients had OPSCC localized in the posterior third (base) of the tongue ($n = 13$), the soft palate ($n = 2$), the palatine tonsils ($n = 13$), and the posterior oropharyngeal wall ($n = 1$).

Imaging
Of the 101 CECT studies included, 72 were performed in our department, and 29, by an external institution. In our department, we perform neck CECT scans on an 80-section CT scanner (Aquilion PRIME; Canon Medical Systems, Otawara, Japan). Our standard protocol includes scout-based automated selection of tube voltages between 100 and 130 kV and tube current modulation between 60 and 600 mA, a tube rotation time of 0.75 seconds, collimated section thickness of 0.5 mm, and a pitch factor of 0.813. Seventy-five milliliters of contrast medium (iopamidol, Imeron 400; Bracco, Milan, Italy) is injected as a split bolus: the first bolus of 50 mL at a flow rate of 2.5 mL/s and the second bolus of 25 mL 55 seconds later at a flow rate of 3.5 mL/s, followed by a 40 mL saline chaser at a flow rate of 2.5 mL/s. The helical scan starts with a delay of 18 seconds after the start of the second bolus injection.

Image quality of the CECT datasets was rated on a 4-point scale (1, excellent; 2, good; 3, acceptable; 4, not acceptable) to ensure that the dataset allows adequate assessment of the primary site, which is often and primarily affected by metal artifacts. A rating of 4 means that the primary site cannot be evaluated for cancer recurrence.

Inclusion Criteria

1) Status posttreatment of OCSCC or OPSCC and recorded case discussion in our weekly interdisciplinary conference (departments of radiology and of oral and maxillofacial surgery).
2) CECT within 3–12 months after treatment or prior surveillance imaging.
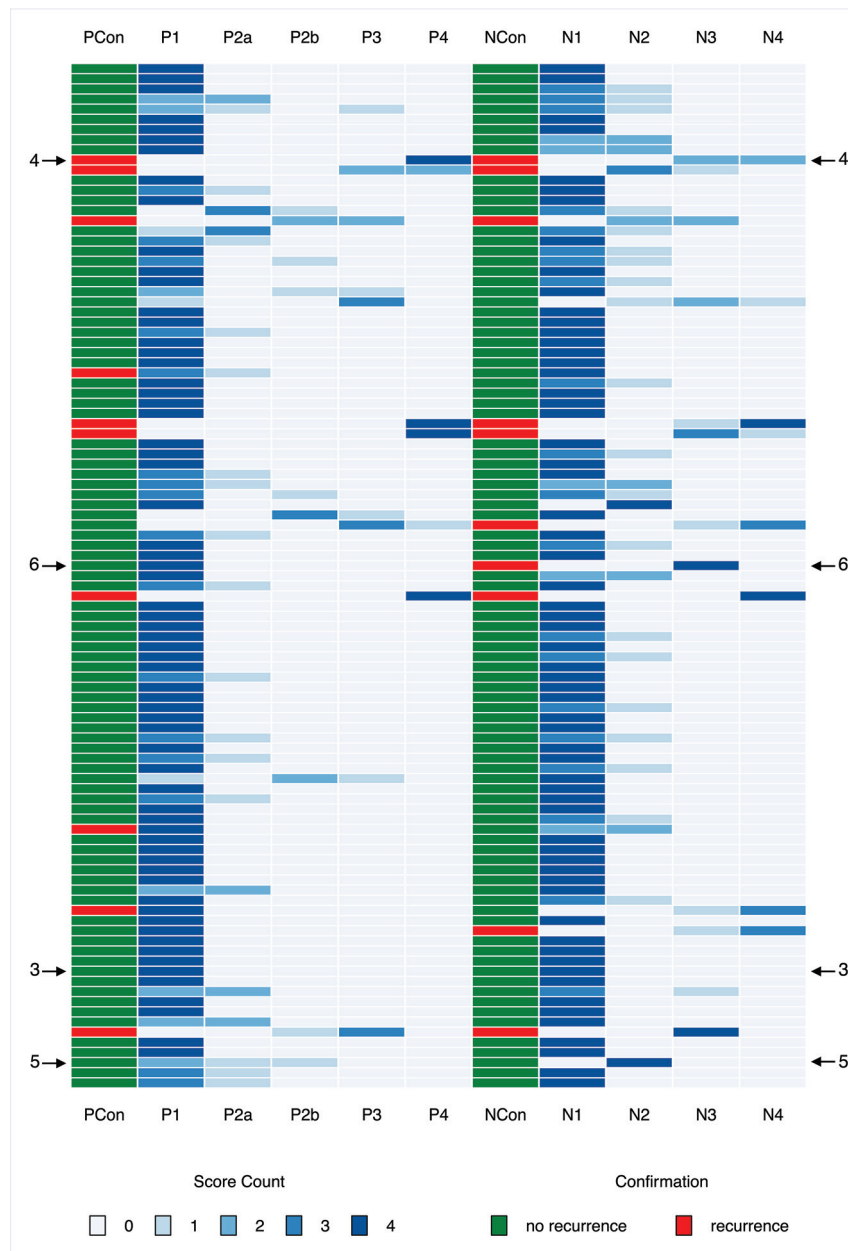3) CECT imaging-quality requirements.

FIG 2. Score distribution chart for all 101 patients. Score counts are coded as shades of blue. Two columns (PCon and NCon) provide the result of the confirmation study. *Arrows* with numbers refer to figures providing CT images of respective patients. PCon and NCon indicate the results of the confirmation studies for the primary site and the neck; P1, P2a, P2b, P3, P4, NI-RADS categories for the primary site; N1, N2, N3, N4, NI-RADS categories for the neck.

a) Split bolus injection of contrast medium resulting in a combined vascular and delayed phase in 1 acquisition.
b) Arms positioned below the head and neck (next to the chest and abdomen).
c) Image quality rating of 1 (excellent), 2 (good), or 3 (acceptable).
4) Confirmation study either as:
a) Subsequent surveillance imaging (CECT, MRI, PET) no earlier than 3 months after the CECT study included or
b) Histopathologic study.

Exclusion Criteria

1) Failure to meet CECT quality requirements:

a) Single bolus injection of contrast medium resulting in a single delayed phase.
b) Arms positioned over the head.
c) Image quality rating of 4 (not acceptable).
2) No subsequent confirmation study.

Readers and Reporting Process

Four radiologists with different levels of experience (A, 3 years and ~300 prior reports of neck CECT; B, 4 years and ~300 reports of neck CECT; C, 7 years and ~700 reports of neck CECT; D, 15 years and ~3300 reports of neck CECT) reviewed the 101 cases included in our analysis. Radiologists A and B were grouped as less
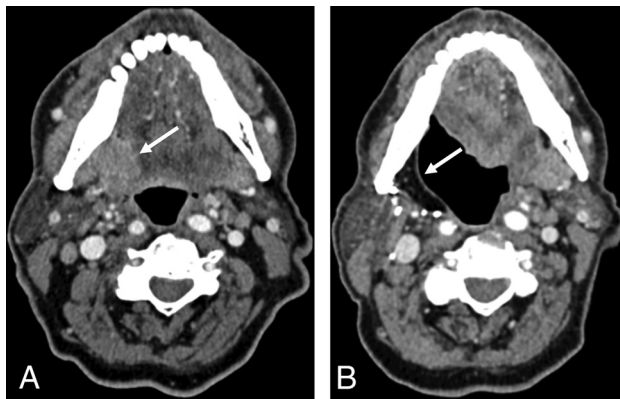
FIG 3. Pretreatment CT (*A*) of a patient with OCSCC located in the right glossopharyngeal sulcus. Posttreatment CT (*B*) of the same patient obtained 36 months after resection and neck dissection on the right side (*B*). A NI-RADS score of 1 was assigned in *B* for the primary site by all 4 readers. The *white arrow* indicates the cancer lesion in the primary site (*A*) and the fattily degenerated muscle flap after resection (*B*).
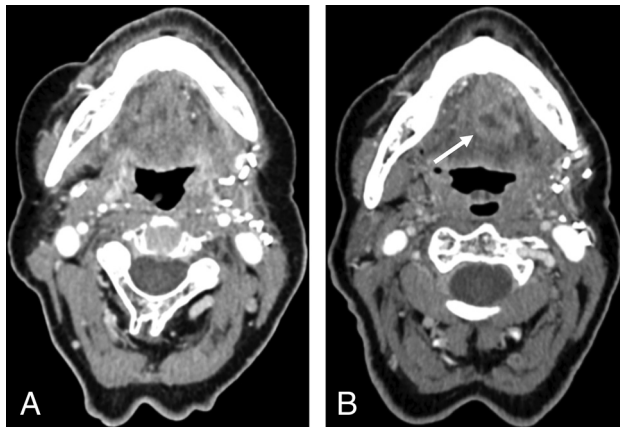

FIG 5. Posttreatment CTs of a patient with OCSCC located in the anterior mouth floor obtained 12 (*A*) and 24 (*B*) months after resection and bilateral neck dissection. The patient's position differed slightly between the 2 posttreatment CT scans. NI-RADS scores of 2a, 2b, 1, and 1 reflect inconsistent interpretation of the primary site (indicated by the *white arrows*) in *B*. Histopathology revealed no malignancy.

FIG 4. Posttreatment CTs of a patient with OPSCC located in the left mouth floor obtained 3 months (*A*) and 15 months (*B*) after resection and neck dissection on the left side. A NI-RADS score of 4 was assigned in *B* for the primary site by all 4 readers. Histopathology confirmed recurrence. The *white arrow* indicates a new enhancing mass in the mouth floor.

FIG 6. Pretreatment CT (*A*) of a patient with OCSCC located in the buccal mucosa in the upper left quadrant. Posttreatment CT (*B*) of the same patient 3 months after resection and neck dissection on the left side shows an enlarged and necrotic parotid lymph node on the left side as indicated by the *white arrows*. A NI-RADS score of 3 was assigned in *B* for the neck by all 4 readers, and histopathology confirmed malignancy.

experienced; C and D, as more experienced readers. Radiologist D is specialized in imaging of the head and neck. At no time were any of the 4 radiologists involved in the interdisciplinary conferences from which patients were included in this study. Anonymized patients were reordered using random numbers assigned by Excel (Version 16.16.10; Microsoft, Redmond, Washington). Readers had access to previous imaging studies (before and after treatment, if available), and they were aware of clinical information to simulate a real reporting situation. Subsequent imaging findings, diagnoses, or clinical examination reports were not available to the 4 readers. After 3 months, radiologists A, B, C, and D were asked again to report on the CECT datasets of the same 101 patients now presented in a newly randomized order. Each of the 2 serial rating sessions was performed in 4 rounds with 25, 25, 25, and 26 patients and a break of 1 week between each round. Another radiologist who was not part of the NI-RADS reader group (E, 6 years of
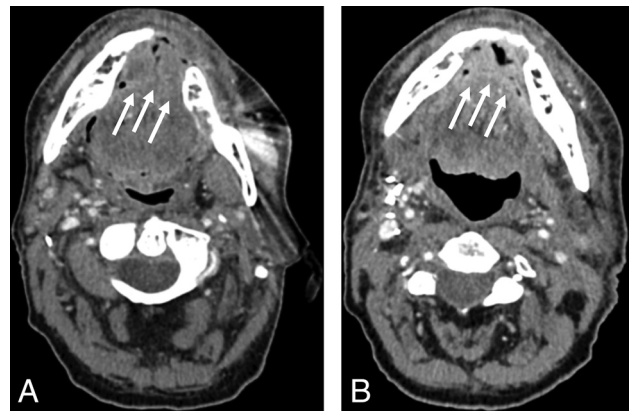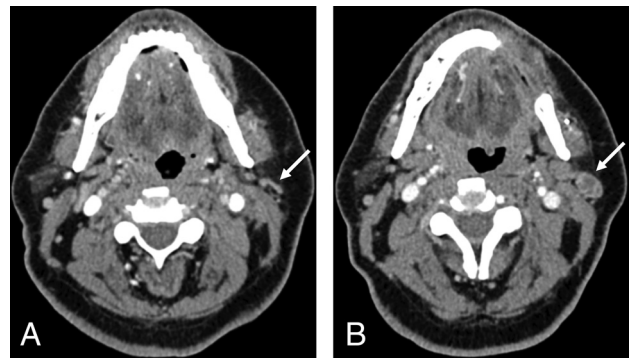
experience and ~400 CECT examinations of the neck) rated the image quality.

NI-RADS Scoring System

Reports of imaging findings were based on the NI-RADS White Paper published in 2018, which was well-studied and jointly discussed by our readers and the authors.[11] NI-RADS scores between 1 and 4, reflecting increasing probabilities of cancer recurrence, are assigned separately for the primary site and for cervical lymph nodes ("neck"). NI-RADS 0 is only used as a preliminary score in cases in which prior images have been obtained but are not available at the time of reading and therefore were not required in our study design. NI-RADS 1 is assigned for expected posttherapeutic changes like the typical superficial diffuse linear contrast enhancement in the primary site and absence of residual abnormal, new, or

Table 1: Interreader agreement

Group	Primary Site				Neck			
	W	τ_B	κ_F	κ_w	W	τ_B	κ_F	κ_w
Overall	0.74		0.48		0.80		0.50	
More experienced readers		0.50		0.80		0.60		0.76
Less experienced readers		0.82		0.96		0.96		0.99
No recurrence	0.56		0.30		0.56		0.29	
Recurrence	0.96		0.65		0.78		0.41	
NI-RADS 1 and 2 merged	0.51		0.23		0.56		0.26	
NI-RADS 3 and 4 merged	0.85		0.56		0.59		0.44	

Table 2: Intrareader agreement

Reader	Primary Site		Neck	
	τ_B	κ_w	τ_B	κ_w
A	0.73	0.85	0.86	0.93
B	0.67	0.90	0.86	0.95
C	0.82	0.96	0.86	0.94
D	0.76	0.92	0.76	0.89

enlarged lymph nodes in the neck. NI-RADS 2 for the primary site is subdivided into 2a for focal superficial enhancement and 2b for deep, ill-defined enhancement. NI-RADS 2 for the neck indicates residual abnormal or new, enlarged lymph nodes without new necrosis or extranodal extension. NI-RADS 3 is assigned for discrete masses in the primary site and new necrosis or extranodal extension of lymph node involvement in the neck. NI-RADS 4 indicates definitive primary site or nodal radiologically or even histopathologically proved recurrence.

Data Analysis

Statistical analysis was performed using R Studio (Version 1.1.383; http://rstudio.org/download/desktop) with the "irr" package installed. The heatmap (Fig 2) was generated using R Studio and the "gplots" package. The flowchart was issued using draw.io (Version 10.8.0; JGraph, Northampton, UK).

Subgroups were formed according to readers' experience (more-versus-less experienced), the results of the confirmation studies (no recurrence versus recurrence), and the probability of cancer recurrence based on the NI-RADS scores of most readers (NI-RADS 1 and 2 versus NI-RADS 3 and 4).

The Kendall's W (W) and Fleiss κ (κ_F) were calculated to test interreader agreement. Calculation of W included a correction factor for tied ranks, and its statistical significance was assessed using the χ^2 test. The Kendall's rank correlation coefficient τ_B and the Cohen weighted κ (κ_w) were computed to quantify either interreader agreement between 2 readers or intrareader agreement. Calculation of κ_w provided weighted disagreements according to their squared distance from perfect agreement.

W and τ_B were interpreted on the basis of the guidelines of Schmidt,[14] proposing a 5-step classification: 0.10–0.29, very weak agreement; 0.30–0.49, weak agreement; 0.50–0.69, moderate agreement; 0.70–0.89, strong agreement; 0.90–1.00, very strong agreement. Interpretation of κ_F and κ_w followed the recommendations of Landis and Koch:[15] < 0.20, slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; 0.81–1.00, (almost) perfect agreement.

Recurrence rates were calculated from the NI-RADS scores of most readers. In case of tied scores, the score assigned by the most experienced reader D was decisive.

RESULTS

Figure 2 provides an overview of rating distributions for all 101 patients in the form of a heatmap. It also includes results of the confirmation studies with arrows indicating exemplary cases with perfect or poor agreement among raters. Numbers next to the arrows indicate the figure in which the cases are presented (Figs 3–6).

Depending on the statistical tests used, overall interreader agreement (Table 1) was strong or moderate for both the primary site (W = 0.74, κ_F = 0.48) and the neck (W = 0.80, κ_F = 0.50). Less experienced readers showed higher interreader agreement for the primary site (τ_B = 0.82 versus 0.50, κ_w = 0.96 versus 0.80) and the neck (τ_B = 0.96 versus 0.60, κ_w = 0.99 versus 0.76). Other subgroups were formed according to the results of the confirmation studies. A total of 13 patients were diagnosed with cancer recurrence. Seven patients had simultaneous cancer recurrence at the primary site and in the neck, while 3 patients each had cancer recurrence at the primary site or in the neck. In patients without proved recurrence, interreader agreement was moderate or fair for the primary site (W = 0.56, κ_F = 0.30) and the neck (W = 0.56, κ_F = 0.29). By contrast, interreader agreement in patients with proved recurrence was very strong or substantial for the primary site (W = 0.96, κ_F = 0.65) and strong or moderate for the neck (W = 0.78, κ_F = 0.41). When forming merged NI-RADS categories according to high and low suspicion of cancer recurrence, we found higher interreader agreement for NI-RADS 3/4 than NI-RADS 1/2 for both the primary site (W = 0.85 versus 0.51, κ_F = 0.56 versus 0.23) and the neck (W = 0.59 versus 0.56, κ_F = 0.44 versus 0.26).

Intrareader agreement (Table 2) for the primary site ranged from moderate to strong (τ_B = 0.67–0.82) or almost perfect (κ_w = 0.85–0.96). Intrareader agreement for the neck was strong (τ_B = 0.76–0.86) or almost perfect (κ_w = 0.89–0.95).

All statistical analyses conducted to test inter- and intrareader agreement showed statistical significance ($P < .05$).

Recurrence rates (Table 3) were between 3.57% (NI-RADS 1) and 100% (NI-RADS 4) for the primary site and 0% (NI-RADS 1) and 83.33% (NI-RADS 4) for lymph nodes (Table 3). Patients without histopathology for confirmation of their diagnosis were followed up for a median of 351 days (range, 159–772 days), defined by the date of their last surveillance imaging study.

DISCUSSION

Inter- and intrareader agreement is important for estimating the reliability of any diagnostic test. To the best of our knowledge, a study investigating inter- and intrareader agreement of NI-RADS scores has not been published. However, we can discuss our results for NI-RADS with those other investigators' results obtained for the reliability of RADS in other organs. Published data give a very diverse picture. A study similar to ours in terms

Table 3: Score counts and recurrence rates for each category based on majority decision[a]

NI-RADS Score	Score Count	Recurrence	No Recurrence	Recurrence Rate
Primary 1	84	3	81	3.57%
Primary 2a	6	0	6	0%
Primary 2b	3	1	2	33.33%
Primary 3	4	2	2	50%
Primary 4	4	4	0	100%
Node 1	83	0	83	0%
Node 2	7	1	6	14.29%
Node 3	5	4	1	80%
Node 4	6	5	1	83.33%

[a] The most experienced reader D was decisive in case of tied score counts.

of statistical methods and results was published by Irshad et al,[16] who assessed consecutive versions of BI-RADS including 5 readers and 104 mammographic examinations. They found an overall interreader agreement of 0.65 and 0.57 (Fleiss κ), while overall intrareader agreement was 0.84 and 0.78 (Cohen weighted κ). A study by Smith et al[17] determined the reliability of PI-RADS in the interpretation of multiparametric MR imaging of the prostate, including 4 readers and 102 examinations, again similar to our study design. However, by contrast, they reported an overall interreader agreement of 0.24 (Fleiss κ) and an overall intrareader agreement of 0.43–0.67 (Cohen κ).

When we compared the 2 studies with our results, the difference in overall interreader agreement stood out first. Our results obtained with NI-RADS (κ_F = 0.48 and 0.50) are much better than findings reported by other investigators for PI-RADS but inferior to results achieved with BI-RADS. NI-RADS showed a very high intrareader agreement (κ_w = 0.85–0.96 and κ_w = 0.89–0.95), especially against the poor values obtained in the PI-RADS study. Thus, our results are encouraging because they suggest that there is the potential for improving interreader agreement. Given that the NI-RADS lexicon and decision tree can only be used fully when interpreting PET/CT or PET/MR imaging, we expect that interreader agreement can be considerably improved using either of these modalities. Especially, NI-RADS categories 1 and 2 (2a and 2b) are defined more clearly when additional information on FDG uptake is available.

Apart from our findings regarding absolute overall agreement, our analysis also provides some interesting results regarding the subgroups formed. Unexpectedly, overall interreader agreement for both the primary site and the neck was higher between the 2 less experienced readers than between the 2 more experienced readers. Furthermore, interreader agreement for the absence of recurrence in lymph nodes was poorer than we expected. A possible explanation emerged from discussions with the readers after completion of the study: The definition for assigning a lymph node to NI-RADS 2 is "mildly enlarging without specific morphologically abnormal features such as new necrosis or extracapsular spread," which was perceived as rather vague.[11] Some kind of measurable threshold might significantly increase agreement among raters. Other results of our study suggest adequate sensitivity of NI-RADS. Interreader agreement was significantly higher in cases of proved cancer recurrence compared with patients without recurrence.

Coincidentally low recurrence rates in the group classified as NI-RADS 1 as well as high recurrence rates in groups with NI-RADS scores of 3 and 4 suggest that NI-RADS is a powerful tool for discrimination of patients with a low-versus-high risk of cancer recurrence. No patients assigned scores of 2a for the primary site had cancer recurrence, which might be attributable to the relatively small number of cases or greater variability in the interpretation of findings, as already discussed above. Recurrence rates calculated in our study are based on majority decision but align very well with initially published data.[11,18,19]

While calculation of κ coefficients is by far the most common statistical test to quantify inter- and intrareader agreement,[20,21] there are also more differentiated approaches addressing other aspects of inter- and intrareader agreement.[22] Other investigators primarily recommend κ statistics for testing nominal scaled data.[23,24] From our standpoint, NI-RADS scores should be regarded as ordinal data because rising values represent a rising probability of cancer recurrence. Therefore, the Kendall's coefficient of concordance (used to determine interreader agreement for >2 readers) and the Kendall's rank correlation coefficient (interreader agreement with 2 raters or intrareader agreement) should be most appropriate.[25] When we compared the result pairs of statistical methods in our study, it is apparent that values of W are always higher than those of κ_F but values of τ_B are always lower than those of κ_w, while their relationships stay basically constant. The intraclass correlation is also used to determine inter- and intrareader agreement; however, it should only be used for underlying continuous data. We therefore chose not to calculate intraclass correlation statistics for the discrete data provided by NI-RADS.

This study, although retrospective, was designed to put readers in a real-world clinical reporting situation. This means that the readers had access to information on OCSCC/OPSCC localization as defined by the multidisciplinary cancer conference, surgical and radiotherapeutic procedures, and pre-existing illnesses. This information is available to reporting radiologists in the clinical setting and is important for appropriately and comprehensively interpreting imaging findings and assessing the patient's condition. On the other hand, there were actions to reduce possible bias. Cases were presented in randomized order, and anonymization of patient data was performed to lower a possible detection bias. The 101 CECT datasets were split into 4 rating sessions (25, 25, 25, and 26) to minimize possible over- or underratings because of readers' raised awareness and altered perception of similarities and differences when comparing cases with others they have recently seen in the artificial reading situation.

Clinically suspected OCSCC or OPSCC and posttherapeutic surveillance are the most frequent indications for neck imaging in our institution, with CECT being much more commonly used than MR imaging. Future studies should investigate inter- and intrareader agreement of NI-RADS, not only for other malignancies (eg, larynx and salivary glands) but also for different imaging modalities (CECT, MR imaging, PET/CT and PET/MR imaging). The role of PET/CT and PET/MR imaging in up- or downgrading lesions seen on CECT or MR imaging without PET should also be of interest in studies, especially prospectively designed, studies.

Limitations

Four radiologists reported imaging findings in this study. While radiologists A and B were relatively close in terms of work experience (years and number of examinations), C and D were wider apart. Although C could easily be classified as more experienced than A and B, a work experience closer to D would have been desirable to ensure ideally balanced subgroups. Subdividing readers into 3 groups with an additional group of intermediate experience might also yield interesting additional results. Because we just started to integrate NI-RADS as a reporting system in our institution, future studies could address these limitations. As readers become more familiar with using NI-RADS and shared experience grows, common approaches might emerge and improve interreader agreement. Although all 4 radiologists were well-acquainted with the literature on NI-RADS, a joint discussion of exemplar cases from our department might have improved interreader and even intrareader agreement. Beyond that, in our opinion, more experience might also lead to higher rates of NI-RADS 2a/b scores being assigned because findings in this category are more difficult to express in prosaic reports because referring clinicians expect a clear decision between "suspected recurrence" versus "no suspected recurrence." We determined recurrence rate as a secondary outcome. Although it attests to the good discriminatory power of NI-RADS, future studies investigating the validity of NI-RADS should define a longer follow-up period of at least 1 year.

CONCLUSIONS

NI-RADS used for interpreting CECT after treatment of OCSCC and OPSCC provides acceptable score reproducibility. A major strength of this standardized approach is the good interreader agreement in patients with proved cancer recurrence and overall intrareader agreement in general. At the same time, there are limitations in terms of interreader agreement in patients with posttherapeutic changes but no cancer recurrence. Although only determined as secondary outcomes, recurrence rates in our patients were similar to those in preliminary published data.

REFERENCES

1. Chi AC, Day TA, Neville BW. **Oral cavity and oropharyngeal squamous cell carcinoma: an update.** *CA Cancer J Clin* 2015;65:401–21 CrossRef Medline
2. Chaturvedi AK, Engels EA, Pfeiffer RM, et al. **Human papillomavirus and rising oropharyngeal cancer incidence in the United States.** *J Clin Oncol* 2011;29:4294–301 CrossRef Medline
3. Bray F, Ferlay J, Soerjomataram I, et al. **Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.** *CA Cancer J Clin* 2018;68:394–424 CrossRef Medline
4. Rivera C. **Essentials of oral cancer.** *Int J Clin Exp Pathol* 2015;8:11884–94 Medline
5. Seeburg DP, Baer AH, Aygun N. **Imaging of patients with head and neck cancer: from staging to surveillance.** *Oral Maxillofac Surg Clin North Am* 2018;30:421–33 CrossRef Medline
6. Queiroz MA, Hullner M, Kuhn F, et al. **PET/MRI and PET/CT in follow-up of head and neck cancer patients.** *Eur J Nucl Med Mol Imaging* 2014;41:1066–75 CrossRef Medline
7. Pfister DG, Spencer S, Brizel DM, et al; National Comprehensive Cancer Network. **Head and neck cancers, Version 2.2014: clinical practice guidelines in oncology.** *J Natl Compr Canc Netw* 2014;12:1454–87 CrossRef Medline
8. D'Orsi CJ, Kopans DB. **Mammography interpretation: the BI-RADS method.** *Am Fam Physician* 1997;55:1548–50, 52 Medline
9. Barentsz JO, Richenberg J, Clements R, et al. **ESUR prostate MR guidelines 2012.** *Eur Radiol* 2012;22:746–57 CrossRef Medline
10. Purysko AS, Remer EM, Coppa CP, et al. **LI-RADS: a case-based review of the new categorization of liver findings in patients with end-stage liver disease.** *Radiographics* 2012;32:1977–95 CrossRef Medline
11. Aiken AH, Rath TJ, Anzai Y, et al. **ACR Neck Imaging Reporting and Data Systems (NI-RADS): A White Paper of the ACR NI-RADS Committee.** *J Am Coll Radiology* 2018;15:1097–108 CrossRef Medline
12. Krieger DA, Hudgins PA, Nayak GK, et al. **Initial performance of NI-RADS to predict residual or recurrent head and neck squamous cell carcinoma.** *AJNR Am J Neuroradiol* 2017;38:1193–99 CrossRef Medline
13. Aiken AH, Farley A, Baugnon KL, et al. **Implementation of a novel surveillance template for head and neck cancer: Neck Imaging Reporting and Data System (NI-RADS).** *J Am Coll Radiol* 2016;13:743–46.e1 CrossRef Medline
14. Schmidt RC. **Managing Delphi surveys using nonparametric statistical techniques.** *Decision Sciences* 1997;28:763–74 CrossRef
15. Landis JR, Koch GG. **The measurement of observer agreement for categorical data.** *Biometrics* 1977;33:159–74 CrossRef
16. Irshad A, Leddy R, Ackerman S, et al. **Effects of changes in BI-RADS Density Assessment Guidelines (fourth versus fifth edition) on breast density assessment: intra- and interreader agreements and density distribution.** *AJR Am J Roentgenol* 2016;207:1366–71 CrossRef Medline
17. Smith CP, Harmon SA, Barrett T, et al. **Intra- and interreader reproducibility of PI-RADSv2: a multireader study.** *J Magn Reson Imaging* 2019;49:1694–703 CrossRef Medline
18. Hsu D, Chokshi FH, Hudgins PA, et al. **Predictive value of first posttreatment imaging using standardized reporting in head and neck cancer.** *Otolaryngol Head Neck Surg* 2019;161:978–85 CrossRef Medline
19. Wangaryattawanich P, Branstetter BF, Hughes M, et al. **Negative predictive value of NI-category 2 in the first posttreatment FDG-PET/CT in head and neck squamous cell carcinoma.** *AJNR Am J Neuroradiol* 2018;39:1884–88 CrossRef Medline
20. Grimm LJ, Anderson AL, Baker JA, et al. **Interobserver variability between breast imagers using the fifth edition of the BI-RADS MRI lexicon.** *AJR Am J Roentgenol* 2015;204:1120–24 CrossRef Medline
21. Rosenkrantz AB, Ginocchio LA, Cornfeld D, et al. **Interobserver reproducibility of the PI-RADS Version 2 Lexicon: a multicenter study of six experienced prostate radiologists.** *Radiology* 2016;280:793–804 CrossRef Medline
22. Griessenauer CJ, Miller JH, Agee BS, et al. **Observer reliability of arteriovenous malformations grading scales using current imaging modalities.** *J Neurosurg* 2014;120:1179–87 CrossRef Medline
23. Hripcsak G, Heitjan DF. **Measuring agreement in medical informatics reliability studies.** *J Biomed Inform* 2002;35:99–110 CrossRef
24. Zapf A, Castell S, Morawietz L, et al. **Measuring inter-rater reliability for nominal data - which coefficients and confidence intervals are appropriate?** *BMC Med Res Methodol* 2016;16:93 CrossRef Medline
25. Kendall MG, Babington Smith B. **The problem of m rankings.** *The Annals of Mathematical Statistics* 1939;10:275–87 CrossRef