Journal of Nuclear Medicine, published on May 12, 2022 as doi:10.2967/jnumed.122.263902

Title:
^{18}F FDG PET/CT Staging of Head and Neck Cancer: Inter-observer agreement and Accuracy – Results from multicenter ACRIN 6685 Clinical Trial

Authors:

Subramaniam RM[1,2]
Duan F[3]
Romanoff J[3]
Yu, JQ[4]
Bartel, T[5]
Dehdashti, F[6]
Intenso, CM[7]
Solnes, L[8]
Sicks, J[3]
Stack BC Jr[9]
Lowe VJ[10]

[1]Otago Medical School, University of Otago, New Zealand, [2]Duke University, Durham, [3]School of Public Health, Brown University, Providence, [4]Fox Chase Cancer Center, Philadelphia, [5]Global Advanced Imaging, Little Rock, [6]Washington University School of Medicine, St Louis, [7]Thomas Jefferson University, Philadelphia, [8]Johns Hopkins School of Medicine, Baltimore, [9]Southern Illinois School of Medicine, Springfield, [10]Mayo Clinic, Rochester.

Disclosure/Support
ACRIN 6685 was supported by the National Cancer Institute through Grants No. U01 CA079778, U01 CA080098, CA180820, and CA180794.

No potential conflict of interests relevant to this article exist.

ClinicalTrials.gov Identifier: NCT00983697

Corresponding Author

Subramaniam RM
Otago Medical School,
201 Great King Street,
Dunedin 9016,
New Zealand.

E mail: rathan.subramaniam@otago.ac.nz

Abstract:

No prior multicenter clinical trial reported inter-observer agreement of FDG PET/CT scans for staging of clinical N0 neck.

Methods: A total of 287 participants were recruited. For visual analysis, "positive" nodal uptake of FDG was defined as uptake visually greater than activity seen in the blood pool.

Results: The Negative Predictive Value of the FDG PET/CT for N0 clinical neck was 86% or above for visual assessment (86% - 88%) for the two central readers and above 90% (90% - 95%) for SUVmax 1.8 and 3.5 cutpoints for central readers and site reads. The kappa coefficients between (1) the two expert readers and (2) between central reads and site reads varied between 0.53 and 0.78.

Conclusion: The NPV of the FDG PET/CT for N0 clinical neck was 86% or above for visual assessment and above 90% for SUVmax 1.8 and 3.5 cutpoints with moderate to substantial agreements.

[^{18}F]fluorodeoxyglucose (FDG) PET/CT is commonly used in clinical practice for management of HNSCC patients including for staging, treatment assessment and detecting recurrence and metastases[1-5]. We previously reported on the primary results of ACRIN 6685 trial [5, 6]. No prior multicenter study reported inter-observer agreement for staging clinical N0 neck in head and neck cancer. In this post-hoc analysis study, we report on the inter-observer agreement among the readers interpreting the FDG PET/CT studies and their accuracy.

Materials and Methods

Patient population

As previously described, a total of 287 participants recruited[5](Figure 1). A clinically N0 neck was defined as being free of palpable lymph nodes and with neck CT and/or MRI neck lymph node sizes of less than 1 cm and 1.5 cm for jugular digastric nodes (IIa), spinal accessory nodes (IIb), or submental-submandibular nodes (Ia and Ib) or showing a lack of central lymph node necrosis in nodes of any size. [5].

Imaging Procedure and Interpretation

Imaging procedures and interpretation methods were previously described in the primary paper and associated supplementary material[5]. PET/CT images were read at each study site by reporting physician (i.e., site reads) and images were presented to a core reading panel of board certified nuclear medicine or nuclear radiology certified physicians. There were two central readers - Reader 1 and Reader 2 (expert head and neck readers) who interpreted most of the PET/CT scans for the study. In addition, Reader 3 and Reader 4 (general readers) were used as central readers 1 and 2 were excluded reading scans from their respective institutions and when there is adjudication needed. A maximum SUV was required for the “hottest” lymph node for

each nodal basin recorded as indeterminate, probably malignant, or definitely malignant. The SUVmax calculation was performed on commercial software (MIM software, version 5.2, Cleveland, OH). For visual analysis, "positive" nodal uptake of FDG was defined as uptake visually greater than background and more than that activity seen in the blood pool (Figure 2).

Statistical Analysis

The neck-level visual assessment FDG-PET/CT scan result for each central reader, for the sites, and for the central adjudicated read was compared to the neck-level pathology result. The sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated. Similar analyses were performed to compare the nodal basin SUVmax result (dichotomized at the optimal cutoff value of 1.8[5] and the pre-specified cutoff value of 3.5) to the nodal-level pathology. Cohen's kappa statistic was used to assess the agreement between: (1) the two expert readers (central readers 1 & 2); (2) the central reads and site reads . Agreement assessment for the two general readers (central readers 3 & 4) was not reported due to data sparsity.

For all analyses, 95% confidence intervals (CIs) were calculated using the 2.5^{th} and 97.5^{th} percentiles of the multilevel bootstrap based on 10,000 resampled datasets[5]. Analyses were performed using SAS software (version 9.4; SAS Institute, Cary, NC) and R (version 4.0.4; R Foundation for Statistical Computing, Vienna, Austria).

Results

Patient demographics

Patient characteristics are included in supplementary table 1S, which include data on enrolled and those who are included in this post-hoc analyses.

Visual Assessment

There were four central readers - Reader 1 and Reader 2 (expert head and neck readers), Reader 3 and Reader 4 (general readers). Readers 1, 2, 3 and 4 interpreted a total of 286, 273, 34, and 26 sides of necks, respectively. The site readers interpreted a total of 296 sides of neck. The sensitivity, specificity, PPV and NPV of the visual assessment for the two expert central readers, the site reads, and the central adjudicated read are summarized in table 1. The kappa coefficients comparing Reader 1 and Reader 2, Reader 1 and the adjudicated central read, Reader 2 and the adjudicated central read, and the site reads and the adjudicated central read were 0.549 (95%CI: 0.431, 0.660), 0.756 (95%CI: 0.664, 0.837), 0.781 (95%CI: 0.696, 0.856), and 0.531 (95%CI: 0.421, 0.633), respectively.

SUVmax Reads

Readers 1, 2, 3 and 4 analyzed a total of 2,272, 2,171, 270, and 208 neck nodes measuring SUVmax, respectively. The site readers analyzed a total of 2,385 neck nodes. The sensitivity, specificity, PPV and NPV of SUVmax for the two expert readers and central adjudicated read are summarized in table 2 for cut-points 1.8 and 3.5. The kappa statistics for measuring the agreement between the site SUVmax and the combined central SUVmax were 0.447 (95%CI: 0.363, 0.527) and 0.525 (95%CI: 0.382, 0.649), respectively for cutpoint SUVmax 1.8 and SUVmax 3.5. The kappa coefficients for measuring the agreement between Reader 1 and the

combined central SUVmax were 0.818 (95%CI: 0.758, 0.870) and 0.751 (95%CI: 0.642, 0.839), respectively for cutpoint SUVmax 1.8 and SUVmax 3.5. The kappa coefficients for measuring the agreement between Reader 2 and the combined central SUVmax were 0.712 (95%CI: 0.640, 0.777) and 0.839 (95%CI: 0.741, 0.915), respectively for cutpoint SUVmax 1.8 and SUVmax 3.5.

Discussion

The NPV of the FDG PET/CT for N0 clinical neck was 86% or above for visual assessment (86% - 88%) for two expert central readers, and above 90% (90% - 95%) for SUVmax 1.8 and 3.5 cutpoints for the two expert readers and site reads. There was moderate to substantial agreement between readers. Increasing evidence supports the higher NPV of PET/CT to exclude nodal metastasis[5, 7-9]. In this study we have provided evidence that multiple readers can achieve high NPV by visual assessment as well as by SUVmax analysis. This result has significant implications, especially managing the contralateral neck, as single center studies have now reported on the outcome of patients managed with observation of PET directed (negative) contralateral neck[10, 11].

The inter reader reliability varied between moderate to substantial agreement in this study. Using the ACRIN 6685 standardized interpretation algorithm (visual assessment) may improve the reliability of interpretation than subjective individual reader interpretation. It is important to note that there was moderate agreement between site readers and central readers, without any training for the site readers, which simulate day to day clinical practice. To our knowledge, there is no other baseline interpretation schema for neck nodal assessment using FDG PET/CT scans

which has undergone inter reader reliability assessment at multicenter level. The standardized qualitative criteria[12] such as Hopkins criteria[2], NI-RADS[13], Deauville[14] and Porceddu[15], are for post therapy settings. The inter-reader reliability for SUVmax readings between central and site readers appears lower than previously reported in single center studies for inter reader and intra reader agreements[16, 17], which is likely due to statistical reporting as a dichotomous (based on cutpoints SUVmax 1.8 and 3.5) measure than continuous measure.

One of the limitations of the ACRIN 6685 reads was no detailed neck nodal level visual interpretation was performed though SUVmax analysis was done. As the visual interpretation was recorded as side of the neck positive or negative for nodal metastasis, a global assessment was obtained. Another limitation for the SUVmax inter- reader agreement is readers may have recorded SUVmax of different lymph node at the same neck nodal level which each reader considered positive and lead to lower inter-reader agreement for SUVmax than observed in single center studies.

In conclusion, the NPV of the FDG PET/CT for N0 clinical neck was 86% or above for visual assessment (86% - 93%) and above 90% (90% - 95%) for SUVmax 1.8 and 3.5 cutpoints. There is moderate to substantial agreement between central readers, between site reads and central adjudicated read, and central readers and central adjudicated read.

Key Points

Question:

What is the negative predictive value (NPV) and reader reliability of FDG PET/CT for staging head and neck cancer with clinical N0 neck in a multicenter trial?

Pertinent Findings:

The NPV of the FDG PET/CT for N0 clinical neck was 86% or above for visual assessment (86% - 88%) and above 90% (90% - 95%) for SUVmax 1.8 and 3.5 cutpoints for the two expert readers and site reads, with moderate to substantial agreement between all readers.

Implication for Patient Care:

FDG PET/CT has very high negative predictive value for staging clinical N0 neck and has moderate to substantial inter-reader reliability, especially between site and central readers, which is important for day to day clinical practice.

References

1. Mehanna, H., et al., *PET-CT Surveillance versus Neck Dissection in Advanced Head and Neck Cancer.* N Engl J Med, 2016. **374**(15): p. 1444-54.
2. Marcus, C., et al., *Head and neck PET/CT: therapy response interpretation criteria (Hopkins Criteria)-interreader reliability, accuracy, and survival outcomes.* J Nucl Med, 2014. **55**(9): p. 1411-6.
3. Van den Wyngaert, T., et al., *Fluorodeoxyglucose-Positron Emission Tomography/Computed Tomography After Concurrent Chemoradiotherapy in Locally Advanced Head-and-Neck Squamous Cell Cancer: The ECLYPS Study.* J Clin Oncol, 2017. **35**(30): p. 3458-3464.
4. Dibble, E.H., et al., *18F-FDG metabolic tumor volume and total glycolytic activity of oral cavity and oropharyngeal squamous cell cancer: adding value to clinical staging.* J Nucl Med, 2012. **53**(5): p. 709-15.
5. Lowe, V.J., et al., *Multicenter Trial of [(18)F]fluorodeoxyglucose Positron Emission Tomography/Computed Tomography Staging of Head and Neck Cancer and Negative Predictive Value and Surgical Impact in the N0 Neck: Results From ACRIN 6685.* J Clin Oncol, 2019. **37**(20): p. 1704-1712.
6. Stack, B.C., Jr., et al., *FDG-PET/CT and Pathology in Newly Diagnosed Head and Neck Cancer: ACRIN 6685 Trial, FDG-PET/CT cN0.* Otolaryngol Head Neck Surg, 2020: p. 194599820969104.
7. Zheng, D., et al., *Relationship between the maximum standardized uptake value of fluoro-2-deoxyglucose-positron emission tomography/computed tomography and clinicopathological characteristics in tongue squamous cell carcinoma.* J Cancer Res Ther, 2019. **15**(4): p. 842-848.
8. Zhao, G., et al., *Significance of PET-CT for Detecting Occult Lymph Node Metastasis and Affecting Prognosis in Early-Stage Tongue Squamous Cell Carcinoma.* Front Oncol, 2020. **10**: p. 386.
9. Linz, C., et al., *Accuracy of 18-F Fluorodeoxyglucose Positron Emission Tomographic/Computed Tomographic Imaging in Primary Staging of Squamous Cell Carcinoma of the Oral Cavity.* JAMA Netw Open, 2021. **4**(4): p. e217083.
10. Zhu, F., S. Sun, and K. Ba, *Comparison Between PET-CT-Guided Neck Dissection and Elective Neck Dissection in cT1-2N0 Tongue Squamous Cell Carcinoma.* Front Oncol, 2020. **10**: p. 720.
11. Hu, K.S., et al., *Low rates of contralateral neck failure in unilaterally treated oropharyngeal squamous cell carcinoma with prospectively defined criteria of lateralization.* Head Neck, 2017. **39**(8): p. 1647-1654.
12. Zhong, J., et al., *Post-treatment FDG PET-CT in head and neck carcinoma: comparative analysis of 4 qualitative interpretative criteria in a large patient cohort.* Sci Rep, 2020. **10**(1): p. 4086.
13. Aiken, A.H., et al., *ACR Neck Imaging Reporting and Data Systems (NI-RADS): A White Paper of the ACR NI-RADS Committee.* J Am Coll Radiol, 2018. **15**(8): p. 1097-1108.

14. Koksel, Y., et al., *Utility of Likert scale (Deauville criteria) in assessment of Chemoradiotherapy response of primary oropharyngeal squamous cell Cancer site.* Clin Imaging, 2019. **55**: p. 89-94.
15. Porceddu, S.V., et al., *Results of a prospective study of positron emission tomography-directed management of residual nodal abnormalities in node-positive head and neck cancer after definitive radiotherapy with or without systemic therapy.* Head Neck, 2011. **33**(12): p. 1675-82.
16. Mhlanga, J.C., et al., *Quantitative PET/CT in clinical practice: assessing the agreement of PET tumor indices using different clinical reading platforms.* Nucl Med Commun, 2018. **39**(2): p. 154-160.
17. Shah, B., et al., *Intra-reader reliability of FDG PET volumetric tumor parameters: effects of primary tumor size and segmentation methods.* Ann Nucl Med, 2012. **26**(9): p. 707-14.

Figure 1: STARD flow diagram

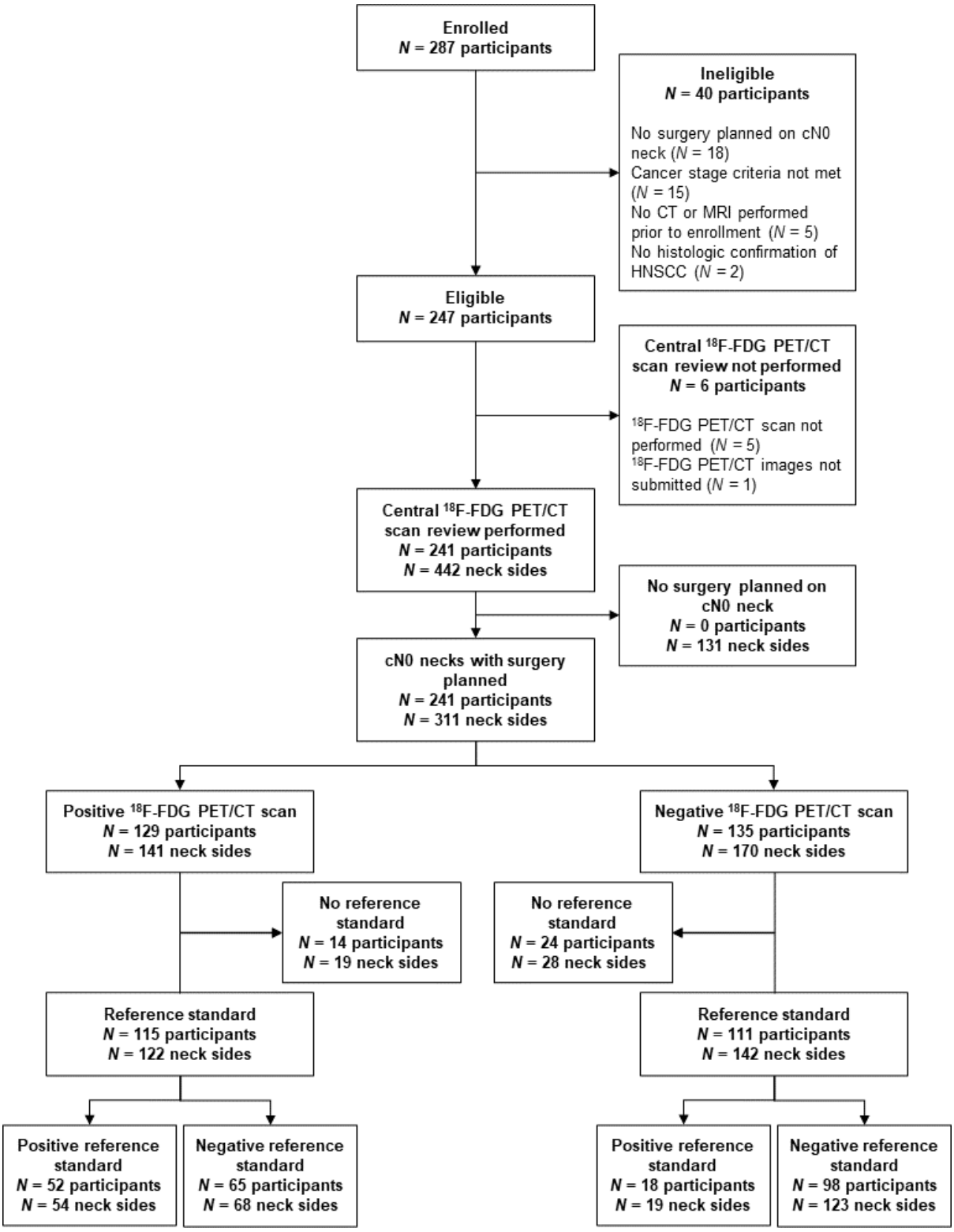

Figure 2: ACRIN 6685 Visual Analysis: Positive and Negative Neck Nodes

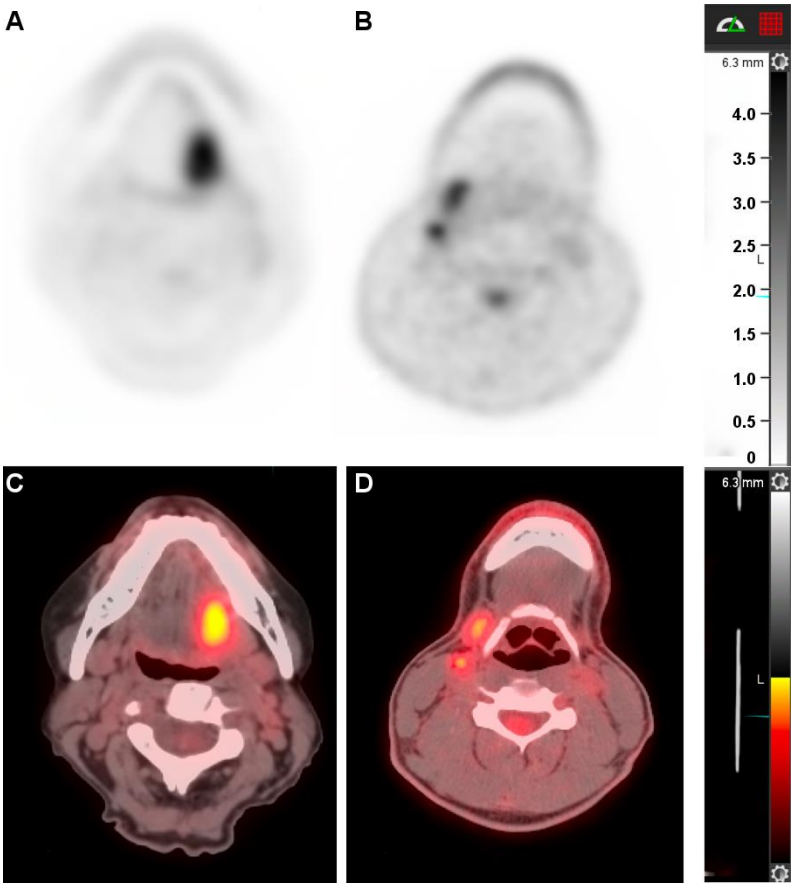

Figure legend: A,C: A negative ^{18}F FDG PET and ^{18}F FDG PET/CT for neck nodes with visual analysis demonstrating FDG uptake in the left level IIA lymph nodes equal or less than FDG uptake in the adjacent blood vessels. SUVmax was 1.1. **B,D:** A positive ^{18}F FDG PET and ^{18}F FDG PET/CT for neck nodes with visual analysis demonstrating FDG uptake in the right level IIA lymph node greater than FDG uptake in the adjacent blood vessels. SUVmax was 3.4.

Table 1: Diagnostic test statistics for the visual assessment FDG-PET/CT scan versus pathology

	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
Expert H&N reader 1	0.791 (0.677, 0.896)	0.584 (0.500, 0.665)	0.417 (0.325, 0.512)	0.881 (0.811, 0.942)
Expert H&N reader 2	0.683 (0.547, 0.810)	0.724 (0.646, 0.797)	0.466 (0.352, 0.583)	0.866 (0.801, 0.925)
Central adjudicated read	0.740 (0.629, 0.845)	0.644 (0.567, 0.716)	0.443 (0.349, 0.538)	0.866 (0.800, 0.924)
Site read	0.700 (0.581, 0.817)	0.699 (0.622, 0.774)	0.471 (0.370, 0.580)	0.859 (0.792, 0.917)

Table 2: Diagnostic test statistics for the dichotomized SUV_{max} result versus pathology

	1.8 cutoff value for SUVmax			
	Sensitivity (95% CI)	**Specificity (95% CI)**	**PPV (95% CI)**	**NPV (95% CI)**
Expert H&N scan reader 1	0.471 (0.327, 0.623)	0.894 (0.862, 0.923)	0.268 (0.167, 0.381)	0.954 (0.931, 0.972)
Expert H&N scan reader 2	0.250 (0.109, 0.419)	0.900 (0.868, 0.929)	0.167 (0.070, 0.281)	0.938 (0.910, 0.962)
Combined central SUV_{max}	0.507 (0.356, 0.652)	0.851 (0.814, 0.884)	0.225 (0.142, 0.315)	0.953 (0.930, 0.972)
Site read	0.395 (0.250, 0.548)	0.903 (0.874, 0.930)	0.263 (0.154, 0.383)	0.945 (0.920, 0.966)
	3.5 cutoff value for SUVmax			
	Sensitivity (95% CI)	**Specificity (95% CI)**	**PPV (95% CI)**	**NPV (95% CI)**
Expert H&N scan reader 1	0.300 (0.155, 0.459)	0.965 (0.942, 0.982)	0.412 (0.231, 0.611)	0.944 (0.919, 0.965)
Expert H&N scan reader 2	0.183 (0.062, 0.330)	0.967 (0.947, 0.983)	0.306 (0.116, 0.517)	0.937 (0.911, 0.961)
Combined central SUV_{max}	0.267 (0.135, 0.412)	0.970 (0.952, 0.986)	0.435 (0.243, 0.634)	0.939 (0.915, 0.961)
Site read	0.250 (0.119, 0.395)	0.972 (0.955, 0.987)	0.442 (0.235, 0.658)	0.937 (0.912, 0.959)

Supplementary Online Table 1S: Baseline participant and disease characteristics

Variable	Eligible participants	
	Primary paper (N=248)	Current report (N=247)
Age, years		
Median (range)	59 (24-95)	59 (24-95)
Race/Ethnicity[1], N (%)		
Asian	35 (14.1)	35 (14.2)
Black or African American	17 (6.9)	17 (6.9)
Hispanic or Latino	4 (1.6)	4 (1.6)
American Indian or Alaskan Native/Unknown	7 (2.8)	7 (2.8)
White	190 (76.6)	189 (76.5)
Sex, N (%)		
Female	84 (33.9)	82 (33.2)
Male	164 (66.1)	165 (66.8)
Primary tumor location, N (%)		
Oral cavity	159 (64.1)	159 (64.4)
Pharynx	43 (17.3)	42 (17.0)
Glottis	47 (19.0)	47 (19.0)
Missing	2 (0.8)	2 (0.8)
Clinical T stage, N (%)		
T1	0 (0.0)	0 (0.0)
T2	153 (61.7)	152 (61.5)
T3	44 (17.7)	44 (17.8)
T4	49 (19.8)	49 (19.8)
Missing	2 (0.8)	2 (0.8)
Clinical N stage, N (%)		
N0	218 (87.9)	216 (87.4)
N1	15 (6.0)	14 (5.7)
N2a	3 (1.2)	4 (1.6)
N2b	8 (3.2)	8 (3.2)
N3	1 (0.4)	2 (0.8)

Variable	Eligible participants	
	Primary paper (N=248)	Current report (N=247)
NX	1 (0.4)	1 (0.4)
Missing	2 (0.8)	2 (0.8)
M0	189 (76.2)	188 (76.1)
MX	57 (23.0)	57 (23.1)
Missing	2 (0.8)	2 (0.8)
Lateralization of tumor		
Right	106 (42.7)	106 (42.9)
Left	108 (43.5)	107 (43.3)
Bilateral	14 (5.6)	14 (5.7)
Midline	18 (7.3)	18 (7.3)
Missing	2 (0.8)	2 (0.8)
Side of neck that was clinically N0		
Right	22 (8.9)	23 (9.3)
Left	18 (7.3)	17 (6.9)
Both sides	206 (83.1)	205 (83.0)
Neither side	0 (0.0)	0 (0.0)
Missing	2 (0.8)	2 (0.8)