# $^{18}$F-FDG PET/CT Staging of Head and Neck Cancer: Interobserver Agreement and Accuracy—Results from Multicenter ACRIN 6685 Clinical Trial

Rathan M. Subramaniam[1,2], Fenghai M. Duan[3], Justin Romanoff[3], Jian Qin Yu[4], Twyla Bartel[5], Farrokh Dehdashti[6], Charles M. Intenzo[7], Lilja Solnes[8], JoRean Sicks[3], Brendan C. Stack Jr[9], and Val J. Lowe[10]

*[1]Otago Medical School, University of Otago, Otago, Dunedin, New Zealand; [2]Duke University, Durham, North Carolina; [3]School of Public Health, Brown University, Providence, Rhode Island; [4]Fox Chase Cancer Center, Philadelphia, Pennsylvania; [5]Global Advanced Imaging, Little Rock, Arkansas; [6]Washington University School of Medicine, St Louis, Missouri; [7]Thomas Jefferson University, Philadelphia, Pennsylvania; [8]Johns Hopkins School of Medicine, Baltimore, Maryland; [9]Southern Illinois School of Medicine, Springfield, Illinois; and [10]Mayo Clinic, Rochester, Minnesota*

To our knowledge, no prior multicenter clinical trial has reported interobserver agreement of $^{18}$F-FDG PET/CT scans for staging of clinical N0 neck in head and neck cancer. **Methods:** A total of 287 participants were recruited. For visual analysis, positive nodal uptake of $^{18}$F-FDG was defined as uptake visually greater than activity seen in the blood pool. **Results:** The negative predictive value of the $^{18}$F-FDG PET/CT for N0 clinical neck was 86% or above for visual assessment (95% CI, 86%–88%) for the 2 central readers and above 90% (95% CI, 90%–95%) for $SUV_{max}$ for central reads and site reads dichotomized at the optimal cutoff value of 1.8 and the prespecified cutoff value of 3.5, respectively. The κ coefficients between the 2 expert readers and between central reads and site reads varied between 0.53 and 0.78. **Conclusion:** The NPV of the $^{18}$F-FDG PET/CT for N0 clinical neck was 86% or above for visual assessment and above 90% for $SUV_{max}$ cut points of 1.8 and 3.5 with moderate to substantial agreements.

**Key Words:** oncology: head and neck; FDG PET/CT; head and neck cancer; staging

**J Nucl Med 2022; 63:1887–1890**
DOI: 10.2967/jnumed.122.263902

PET/CT with $^{18}$F-FDG is commonly used in clinical practice for management of head and neck squamous cell carcinoma patients including for staging, treatment assessment, and detecting recurrence and metastases (*1–5*). We previously reported on the primary results of ACRIN 6685 trial (ClinicalTrials.gov identifier: NCT00983697) (*5,6*). No prior multicenter study reported interobserver agreement for staging clinical N0 neck in head and neck cancer. In this post hoc analysis study, we report on the interobserver agreement among the readers interpreting the $^{18}$F-FDG PET/CT studies and their accuracy.

## MATERIALS AND METHODS

### Patient Population

As previously described, a total of 287 participants were recruited (Fig. 1) (*5*). A clinically N0 neck was defined as being free of palpable lymph nodes and with neck CT or MRI neck lymph node sizes of less than 1 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 (*5*). PET/CT images were read at each study site by the 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 2 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 because central readers 1 and 2 were excluded from reading scans from their respective institutions and when adjudication was needed. A $SUV_{max}$ was required for the hottest lymph node for each nodal basin recorded as indeterminate, probably malignant, or definitely malignant. The $SUV_{max}$ calculation was performed using commercial software (version 5.2; MIM Software). For visual analysis, positive nodal uptake of $^{18}$F-FDG was defined as uptake visually greater than background and more than that activity seen in the blood pool (Fig. 2).

### Statistical Analysis

The neck-level visual assessment $^{18}$F-FDG PET/CT scan result for each central reader, for the sites and for the central adjudicated read, was compared with 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 $SUV_{max}$ result (dichotomized at the optimal cutoff value of 1.8 (*5*) and the prespecified cutoff value of 3.5) with the nodal-level pathology. Cohen's κ statistic was used to assess the agreement between the 2 expert readers (central readers 1 and 2) and the central reads and site reads. Because of data sparsity, agreement assessment for the 2 general readers (central readers 3 and 4) was not reported.

For all analyses, 95% CIs were calculated using the 2.5 and 97.5 percentiles of the multilevel bootstrap based on 10,000 resampled datasets (*5*). Analyses were performed using SAS software (version 9.4; SAS Institute) and R (version 4.0.4; R Foundation for Statistical Computing).

## RESULTS

### Patient Demographics

Patient characteristics are included in Supplemental Table 1 (supplemental materials are available at http://jnm.snmjournals.org),

Received Jan. 25, 2022; revision accepted Apr. 27, 2022.
For correspondence or reprints, contact Rathan M. Subramaniam (rathan.subramaniam@otago.ac.nz).
Published online May 12, 2022.
COPYRIGHT © 2022 by the Society of Nuclear Medicine and Molecular Imaging.

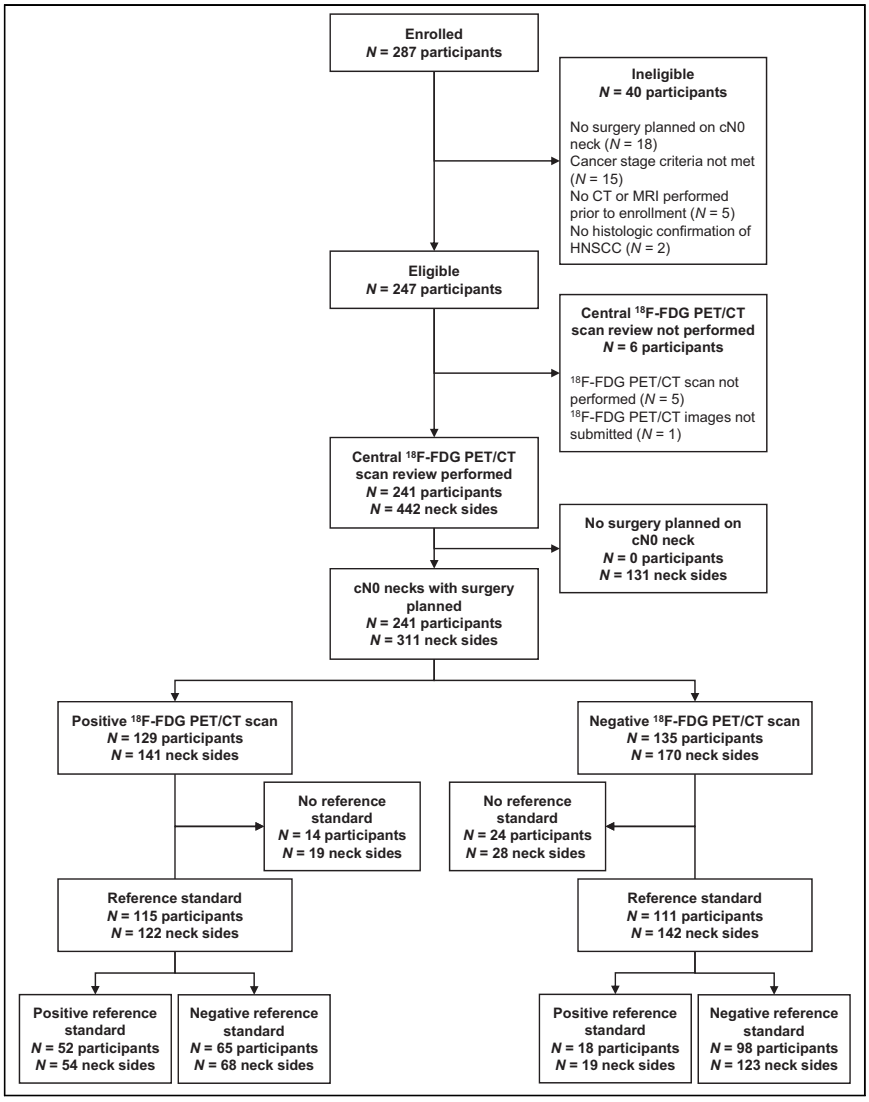


**FIGURE 1.** STARD flow diagram.

which include data on enrolled patients and those who were included in this post hoc analyses.

### Visual Assessment

There were 4 central readers: reader 1 and reader 2 (expert head and neck readers), and 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 2 expert central readers, the site reads, and the central adjudicated read are summarized in Table 1. The κ coefficients comparing reader 1 and reader 2, reader 1 and the central adjudicated read, reader 2 and the central adjudicated read, and the site reads and the central adjudicated 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.

### $SUV_{max}$ Reads

Readers 1, 2, 3, and 4 analyzed a total of 2,272, 2,171, 270, and 208 neck nodes measuring $SUV_{max}$, respectively. The site readers analyzed a total of 2,385 neck nodes. The sensitivity, specificity, PPV, and NPV of $SUV_{max}$ for the 2 expert readers and central adjudicated read are summarized in Table 2 for cut points 1.8 and 3.5. The κ statistics for measuring the agreement between the site $SUV_{max}$ and the combined central $SUV_{max}$ were 0.447 (95% CI: 0.363, 0.527) and 0.525 (95% CI: 0.382, 0.649), respectively, for $SUV_{max}$ cut points of 1.8 and $SUV_{max}$ 3.5. The κ coefficients for measuring the agreement between reader 1 and the combined central $SUV_{max}$ were 0.818 (95% CI: 0.758, 0.870) and 0.751 (95% CI: 0.642, 0.839), respectively, for $SUV_{max}$ cut points of 1.8 and $SUV_{max}$ 3.5. The κ coefficients for measuring the agreement between reader 2 and the combined central $SUV_{max}$ were 0.712 (95% CI: 0.640, 0.777) and 0.839 (95% CI: 0.741, 0.915), respectively, for $SUV_{max}$ cut points 1.8 and 3.5.

## DISCUSSION

The NPV of the $^{18}$F-FDG PET/CT for N0 clinical neck was 86% or above for visual assessment (95% CI, 86%–88%) for 2 expert central readers, and above 90% (95% CI, 90%–95%) for $SUV_{max}$ cut points of 1.8 and 3.5 for the 2 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 $SUV_{max}$ 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 interreader reliability varied between moderate and substantial agreement in this study. Using the ACRIN 6685 standardized interpretation algorithm (visual assessment) may improve the reliability of interpretation more 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 simulates day-to-day clinical practice. To our knowledge, there is no other baseline interpretation schema for neck nodal assessment using $^{18}$F-FDG PET/CT scans that has undergone interreader reliability assessment at a 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 interreader reliability for $SUV_{max}$ readings between central and site readers appears lower than previously reported in single-center studies for interreader and intrareader agreements (*16,17*), which is likely due to statistical reporting as a dichotomous (based on $SUV_{max}$ cut points of 1.8 and 3.5) measure than a continuous measure.

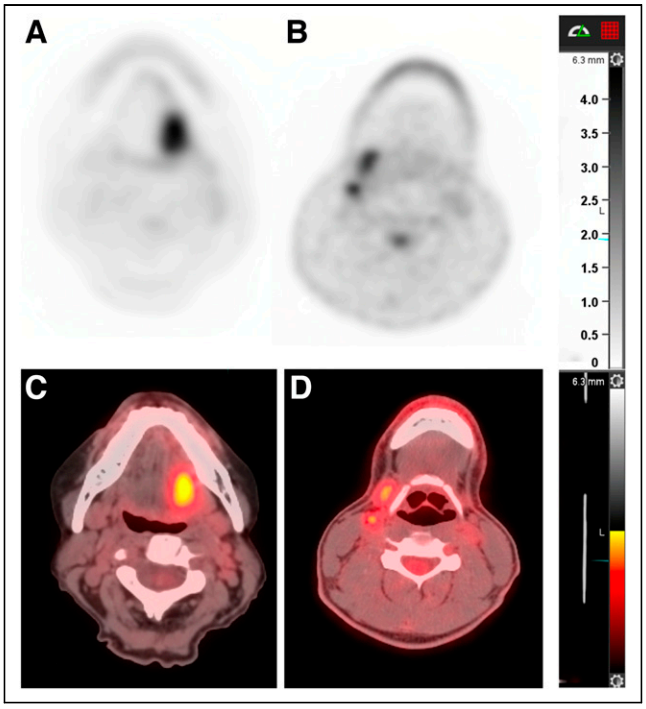


**FIGURE 2.** ACRIN 6685 visual analysis: positive and negative neck nodes. (A and C) Negative $^{18}$F-FDG PET and $^{18}$F-FDG PET/CT findings for neck nodes, with visual analysis demonstrating $^{18}$F-FDG uptake in left level IIA lymph nodes equal to or less than $^{18}$F-FDG uptake in adjacent blood vessels. $SUV_{max}$ was 1.1. (B and D) Positive $^{18}$F-FDG PET and $^{18}$F-FDG PET/CT findings for neck nodes, with visual analysis demonstrating $^{18}$F-FDG uptake in right level IIA lymph node greater than $^{18}$F-FDG uptake in adjacent blood vessels. $SUV_{max}$ was 3.4.

One of the limitations of the ACRIN 6685 reads was that no detailed neck nodal level visual interpretation was performed though $SUV_{max}$ 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 $SUV_{max}$ interreader agreement is readers may have recorded $SUV_{max}$ of different lymph nodes at the same neck nodal level, which each reader considered positive and led to lower interreader agreement for $SUV_{max}$ than observed in single-center studies.

## CONCLUSION

The NPV of the $^{18}$F-FDG PET/CT for N0 clinical neck was 86% or above for visual assessment (95% CI, 86%–88%) and above 90% (95% CI, 90%–95%) for $SUV_{max}$ cut points of 1.8 and 3.5. There is moderate to substantial agreement between central readers, between site reads and central adjudicated read, and central readers and central adjudicated read.

## DISCLOSURE

ACRIN 6685 was supported by the National Cancer Institute through grants U01 CA079778, U01 CA080098, CA180820, and CA180794. No other potential conflict of interest relevant to this article was reported.

## KEY POINTS

**QUESTION:** What is the NPV and reader reliability of $^{18}$F-FDG PET/CT for staging head and neck cancer with clinical N0 neck in a multicenter trial?

**PERTINENT FINDINGS:** The NPV of the $^{18}$F-FDG PET/CT for N0 clinical neck was 86% or above for visual assessment (95% CI, 86%–88%) and above 90% (95% CI, 90%–95%) for $SUV_{max}$ cut points of 1.8 and 3.5 for the 2 expert readers and site reads, with moderate to substantial agreement between all readers.

**IMPLICATIONS FOR PATIENT CARE:** $^{18}$F-FDG PET/CT has very high NPV for staging clinical N0 neck and has moderate to substantial interreader reliability, especially between site and central readers, which is important for day-to-day clinical practice.

## REFERENCES

1. Mehanna H, Wong W-L, McConkey CC, et al. PET-CT surveillance versus neck dissection in advanced head and neck cancer. *N Engl J Med.* 2016;374:1444–1454.
2. Marcus C, Ciarallo A, Tahari AK, et al. Head and neck PET/CT: therapy response interpretation criteria (Hopkins Criteria)-interreader reliability, accuracy, and survival outcomes. *J Nucl Med.* 2014;55:1411–1416.
3. Van den Wyngaert T, Helsen N, Carp L, 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;35:3458–3464.
4. Dibble EH, Lara Alvarez AC, Truong M-T, Mercier G, Cook EF, Subramaniam RM. $^{18}$F-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;53:709–715.
5. Lowe VJ, Duan F, Subramaniam RM, 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;37:1704–1712.

**TABLE 1**
Diagnostic Test Statistics for Visual Assessment $^{18}$F-FDG PET/CT Scan Versus Pathology

| Reader | Sensitivity | Specificity | PPV | NPV |
|---|---|---|---|---|
| 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) |

H&N = head and neck.
Data in parentheses are 95% CIs.

**TABLE 2**
Diagnostic Test Statistics for the Dichotomized $SUV_{max}$ Result Versus Pathology

| Reader | 1.8 cutoff value for $SUV_{max}$ | | | | 3.5 cutoff value for $SUV_{max}$ | | | |
|---|---|---|---|---|---|---|---|---|
| | Sensitivity | Specificity | PPV | NPV | Sensitivity | Specificity | PPV | NPV |
| 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) | 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 | 0.250 (0.109, 0.419) | 0.900 (0.868, 0.929) | 0.167 (0.070, 0.281) | 0.938 (0.910, 0.962) | 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 $SUV_{max}$ | 0.507 (0.356, 0.652) | 0.851 (0.814, 0.884) | 0.225 (0.142, 0.315) | 0.953 (0.930, 0.972) | 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 | 0.395 (0.250, 0.548) | 0.903 (0.874, 0.930) | 0.263 (0.154, 0.383) | 0.945 (0.920, 0.966) | 0.250 (0.119, 0.395) | 0.972 (0.955, 0.987) | 0.442 (0.235, 0.658) | 0.937 (0.912, 0.959) |

H&N = head and neck.
Data in parentheses are 95% CIs.


6. Stack BC Jr, Duan F, Subramaniam RM, 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.* 2021;164:1230–1239.
7. Zheng D, Niu L, Liu W, 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;15:842–848.
8. Zhao G, Sun J, Ba K, Zhang Y. Significance of PET-CT for detecting occult lymph node metastasis and affecting prognosis in early-stage tongue squamous cell carcinoma. *Front Oncol.* 2020;10:386.
9. Linz C, Brands RC, Herterich T, 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:e217083.
10. Zhu F, Sun S, Ba K. Comparison between PET-CT-guided neck dissection and elective neck dissection in cT1-2N0 tongue squamous cell carcinoma. *Front Oncol.* 2020;10:720.
11. Hu KS, Mourad WF, Gamez M, et al. Low rates of contralateral neck failure in unilaterally treated oropharyngeal squamous cell carcinoma with prospectively defined criteria of lateralization. *Head Neck.* 2017;39:1647–1654.
12. Zhong J, Sundersingh M, Dyker K, 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;10:4086.
13. 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 Radiol.* 2018;15:1097–1108.
14. Koksel Y, Gencturk M, Spano A, Reynolds M, Roshan S, Caycı Z. Utility of Likert scale (Deauville criteria) in assessment of chemoradiotherapy response of primary oropharyngeal squamous cell cancer site. *Clin Imaging.* 2019;55:89–94.
15. Porceddu SV, Pryor DI, Burmeister E, 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;33:1675–1682.
16. Mhlanga JC, Chirindel A, Lodge MA, Wahl RL, Subramaniam RM. Quantitative PET/CT in clinical practice: assessing the agreement of PET tumor indices using different clinical reading platforms. *Nucl Med Commun.* 2018;39:154–160.
17. Shah B, Srivastava N, Hirsch AE, Mercier G, Subramaniam RM. Intra-reader reliability of FDG PET volumetric tumor parameters: effects of primary tumor size and segmentation methods. *Ann Nucl Med.* 2012;26:707–714.