Prospective Risk-Adjusted $[^{18}\text{F}]$Fluorodeoxyglucose Positron Emission Tomography and Computed Tomography Assessment of Radiation Response in Head and Neck Cancer


ABSTRACT

Purpose

$[^{18}\text{F}]$Fluorodeoxyglucose positron emission tomography (FDG-PET)/computed tomography (CT) imaging may improve assessment of radiation response in patients with head and neck cancer, but it is not yet known for which patients this is most useful. We conducted a prospective trial to identify patient populations likely to benefit from the addition of functional imaging to the assessment of radiotherapy response.

Patients and Methods

Ninety-eight patients with locally advanced cancer of the oropharynx, larynx, or hypopharynx were prospectively enrolled and treated with primary radiotherapy, with or without chemotherapy. Patients underwent FDG-PET/CT and contrast-enhanced CT imaging 8 weeks after completion of treatment. Functional and anatomic imaging response was correlated with clinical and pathologic response. Imaging accuracy was then compared between imaging modalities.

Results

Although postradiation maximum standard uptake values were significantly higher in nonresponders compared with responders, the positive and negative predictive values of FDG-PET/CT scanning were similar to those for CT alone in the unselected study population. Subset analyses revealed that FDG-PET/CT outperformed CT alone in response assessment for patients at high risk for treatment failure (those with human papillomavirus [HPV]–negative disease, nonoropharyngeal primaries, or history of tobacco use). No benefit to FDG-PET/CT was seen for low-risk patients lacking these features.

Conclusion

These data do not support the broad application of FDG-PET/CT for radiation response assessment in unselected head and neck cancer patients. However, FDG-PET/CT may be the imaging modality of choice for patients with highest risk disease, particularly those with HPV-negative tumors. Optimal timing of FDG-PET/CT imaging after radiotherapy merits further investigation.

J Clin Oncol 27:2509-2515. © 2009 by American Society of Clinical Oncology

INTRODUCTION

Primary radiotherapy, with or without concurrent chemotherapy, for organ-sparing treatment of head and neck squamous cell carcinoma (HNSCC) has advanced markedly over the last decade. However, there remains room for improvement, with local control rates as low as 50% in recent phase III trials. For patients with persistent disease after radiation, consolidative neck dissection and/or surgical salvage of primary disease potentially improve survival. However, such surgery carries significant risk for morbidity and mortality. This underscores the importance of correctly identifying those patients in need of salvage therapy.

The ideal method of treatment response assessment would noninvasively achieve accurate prediction of residual viable tumor after radiotherapy. Clinical examination alone is inadequate for response assessment, providing accuracy less than 50%. Postradiation computed tomography (CT) imaging yields higher sensitivity but has been criticized for lacking specificity.
Recent studies have proposed that functional imaging, specifically \(^{18}\text{F}\)fluorodeoxyglucose positron emission tomography (FDG-PET)/CT, may improve the accuracy of radiation response assessment for HNSCC. These studies are limited by their retrospective design, and consequently, the role of FDG-PET/CT in radiation response assessment remains undefined.

We sought to more definitively address this issue by conducting a prospective trial designed to compare the accuracy of radiation response assessment by FDG-PET/CT and contrast-enhanced CT and define patient subsets likely to derive maximal benefit from the addition of FDG-PET/CT imaging to conventional response assessment.

**Results**

**Patients**

Between November 2005 and May 2007, 107 consecutive patients with locally advanced HNSCC scheduled to receive definitive radiation with or without concurrent systemic therapy were screened for enrollment onto this prospective institutional review board–approved trial. Adult patients were eligible for inclusion if they had biopsy-proven American Joint Commission on Cancer (AJCC) version 6 stage III to IVB squamous cell carcinoma of the oropharynx, hypopharynx, or larynx. Nineteen screened patients were ineligible for inclusion; two withdrew consent, five had disease removed surgically before radiotherapy, and two had FDG-PET/CT scans performed earlier than stipulated by protocol, leaving an enrolled study cohort of 98 patients.

Demographic and clinical covariates were directly tabulated from each patient’s clinical chart. A threshold level of 10 pack-years of use was chosen to distinguish tobacco users from nonusers because this level of exposure is correlated with increased risk of developing HNSCC.

**Radiotherapy and Systemic Treatment**

Patients were treated with intensity-modulated radiotherapy as described previously. Prescription doses ranged from 66 to 70 Gy in 32 to 35 fractions. Intensity-modulated radiotherapy was delivered via a step-and-shoot, multileaf collimation through a static treatment gantry. Treatment planning was performed with a Pinnacle system (version 6.2b or later; Philips Medical Systems, Andover, MA). Seventy-eight patients received concurrent systemic therapy with radiation.

**FDG-PET/CT Imaging**

Two FDG-PET/CT scans were scheduled for each patient; the first was scheduled within 4 weeks before starting therapy, and the second was scheduled as near as possible to the eighth week after completion of treatment. Deviations from the 8-week target were allowed, as long as imaging was performed between 5 and 12 weeks after radiotherapy. Details of our functional imaging protocol are provided in the Appendix (online only). The intensity of metabolic activity within foci of increased FDG uptake in disease sites was analyzed on the FDG-PET images using a semiautomated vendor-provided tool (GE Advantage Workstation; GE Medical Systems, Milwaukee, WI). A single clinician (V.R.) calculated maximum standardized uptake values (SUV\(_{\text{max}}\)) from a volume of interest completely encompassing each site of FDG accumulation. For patients with multiple sites of lymphadenopathy, nodal SUV\(_{\text{max}}\) was calculated on each FDG-PET/CT scan for the node with the highest initial SUV\(_{\text{max}}\). FDG-PET/CT scans were interpreted with knowledge of contrast-enhanced CT scans.

**CT Imaging**

Two contrast-enhanced head and neck CT scans were scheduled for each patient; the first was scheduled before starting therapy, and the second was scheduled as near as possible to the eighth week after treatment. All images were interpreted by a board-certified neuroradiologist (L.E.G.). Interpretation of conventional CT imaging was blinded to FDG-PET/CT results. For further details of the anatomic imaging protocol, refer to the Appendix.

**Post-Treatment Surveillance and Surgical Management**

Patients returned for routine post-treatment surveillance 8 weeks after treatment for baseline assessment and then every 3 to 4 months for the first 2 years for clinical examination and serial contrast-enhanced CT imaging of the head and neck. Chest x-rays were performed annually.

A standardized clinical pathway was followed for postradiotherapy neck management (see Appendix). Consolidative neck dissections were performed at the discretion of the treating physicians and as indicated by findings from all post-treatment studies including, but not limited to, FDG-PET/CT. Patients with no radiographic or clinical evidence for residual adenopathy were observed expectantly. Patients found by clinical or radiographic evaluation to have residual or recurrent primary disease underwent salvage resection.

Histology was the gold standard for defining nonresponse to therapy for patients who underwent surgical consolidation. For patients observed clinically, the gold standard was any locoregional failure detected by physical examination (including fiberoptic pharyngolaryngoscopy) or radiographic imaging (CT, PET/CT, and/or ultrasound) within 6 months of completing radiotherapy.

**Pathologic Tissue Assessment**

Twenty-two patients underwent consolidative neck dissection or salvage resection of primary disease at a median of 85 days after completion of radiotherapy. All neck dissection tissue specimens were evaluated and step sectioned by the same board-certified pathologist (M.D.W.). For further details of the step-sectioning protocol, refer to the Appendix.

**Human Papillomavirus Detection**

When tissue was available, polymerase chain reaction (PCR), in situ hybridization (ISH), and immunohistochemistry were run on paraffin-embedded tumor samples to detect the presence or absence of human papillomavirus (HPV). See the Appendix for further details. Samples were regarded as positive if HPV was detected by either PCR or ISH. Discrepancies between methods were rare and were resolved by p16 immunohistochemistry; positive staining for p16, with HPV DNA detected by either PCR or ISH, was coded as positive for HPV.

**Data Analysis**

Locoregional disease control, distant metastasis-free survival, disease-specific survival, and overall survival from the date of treatment completion were estimated using the Kaplan-Meier method. Patients were separated into two groups, complete responders and nonresponders, based on clinical and pathologic response assessment as described earlier. Multivariate Cox regression analysis (WinStat; http://www.winstat.com/) was used to analyze covariates for correlation with failure (significance defined as \(P < .05\)). SUV\(_{\text{max}}\) values were compared between the groups using a two-tailed \(t\) test (significance defined as \(P < .05\)). Receiver operating characteristic (ROC) curve analysis was performed for the postradiation FDG-PET/CT data sets, separately for primary and nodal disease. Polynomial curves were fit to the ROC data and solved for a slope of 1 to identify SUV\(_{\text{max}}\) thresholds with maximum accuracy for predicting failure. These values were then applied to the raw postradiation FDG-PET/CT data to calculate the accuracy of predicting response (see Appendix).

**Patient Characteristics**

The median follow-up time for all patients enrolled onto this study was 92 weeks (range, 37 to 122 weeks). Two years after the completion of radiotherapy, the predicted overall survival rate for the enrolled cohort was 89.1%, disease-specific survival rate was 90.3%, locoregional control rate was 88.2%, and distant
metastasis-free survival rate was 93.1%. The baseline demographic and clinical characteristics of the enrolled patients were typical for locally advanced HNSCC patients treated at our institution (Appendix Table A1, online only). The majority of patients were nonsmoking white males with stage IV oropharyngeal squamous cell carcinoma treated with 70 Gy over 6 weeks with concurrent weekly cisplatin.

Six patients were censored after experiencing distant failure and initiating systemic therapy before developing locoregional recurrence. This left 92 patients available for further analysis of locoregional disease response. These patients underwent surgical or clinical evaluation of treatment response at the discretion of the treating physician. Eighty patients responded to radiotherapy, and 12 were nonresponders (Appendix Fig A1, online only). Multivariate Cox regression analysis was performed to examine whether baseline clinical or demographic data predicted for response to therapy. Advanced T stage and a history of significant tobacco use were both significant predictors of failure ($P = .045$ and $P = .038$, respectively).

**Correlating SUV<sub>max</sub> With Treatment Response**

We first examined whether SUV<sub>max</sub> values correlated with response to treatment (Fig 1). There were no significant differences between responders and nonresponders in baseline SUV<sub>max</sub> values for primary tumors or nodes (Figs 1A and 1B). However, after radiation, SUV<sub>max</sub> values were significantly higher in nonresponders compared with responders for both primary tumors and nodes ($P < .001$; Figs 1C to 1F). We also examined whether relative or absolute change in SUV<sub>max</sub> from the pre- to postradiation FDG-PET/CT scans correlated with response (Appendix Table A2, online only). The relative change in SUV<sub>max</sub> for primary tumors was the only other parameter significantly associated with response to treatment ($P < .001$).
Accuracy of Post-Treatment SUV\textsubscript{max} for Predicting Treatment Response

Because absolute postradiation SUV\textsubscript{max} was the sole factor correlating with response for both primary tumors and nodal disease, we focused the remainder of our analyses on these data. ROC curve analysis was performed on the postradiation FDG-PET/CT data set to define threshold SUV\textsubscript{max} values with highest accuracy for radiation response assessment (Fig 2). We identified 6.5 and 2.8 as threshold SUV\textsubscript{max} values with maximal accuracy for predicting failure in primary tumors and nodes, respectively. Clinical interpretation of each FDG-PET/CT scan agreed with the results of SUV\textsubscript{max} analysis for all patients on study.

Using these threshold values, the sensitivity, specificity, positive predictive value, and negative predictive value were each calculated for the postradiation FDG-PET/CT data set. For comparison, these values were also calculated for the CT data set, using complete radiographic response as the predictor for clinical cure and all other radiographic end points (partial response, stable disease, and progressive disease) as predictors for failure. Overall, the accuracy of FDG-PET/CT was similar to that of CT (Table 1). When compared with CT alone, FDG-PET/CT had slightly higher specificity and positive predictive value at a cost of lower sensitivity and negative predictive value.

Risk Stratification and FDG-PET/CT Accuracy

The positive predictive value of FDG-PET/CT in our patient cohort was relatively low; it performed better than chance only in predicting residual primary disease and then only by a small margin. The positive predictive value of any test can be improved by limiting its use to patients with a high incidence of events. Therefore, we reasoned that the accuracy of FDG-PET/CT in radiation response assessment could be maximized by applying it to patients at highest risk for treatment failure.

We divided our patients into low-risk and high-risk groups by the following factors correlating with a lack of response to treatment: T stage (0 to 2 \( \leq 3 \) to 4), N stage (0 to 2a \( \leq 3 \) to 2b to 3), overall American Joint Commission on Cancer stage (II to III \( \leq IV \)), primary tumor site (oropharyngeal \( \leq \) nonoropharyngeal), smoking history (user \( \leq \) nonuser), and tumor HPV status \( (positive \leq negative) \). When possible, groupings were designed to divide patients into equal-sized subgroups.

The positive predictive value of FDG-PET/CT did not surpass 50% for any patient subset separated by stage (T, N, or overall). However, this objective was reached for patients with nonoropharyngeal primaries, positive smoking histories, and HPV-negative tumors (Fig 3). The greatest advantage in positive predictive value was seen by dividing patients by tumor HPV status \( (62.5\% \text{ for HPV-negative tumors} \leq 20\% \text{ for HPV-positive tumors}) \), although HPV typing was available for only one third of our patient cohort.

We then constructed a risk stratification algorithm based on the previous data. Patients with HPV-positive tumors or with oropharyngeal primary tumors and no history of tobacco use were classified as low risk \( (n = 61) \). Those with HPV-negative tumors, nonoropharyngeal primaries, or significant tobacco use histories were classified as...

![Fig 2. (18F)Fluorodeoxyglucose positron emission tomography (FDG-PET)/computed tomography (CT) receiver operating characteristic curves are shown (primary tumors —, nodal disease ——). For comparison, the sensitivity/specificity of CT alone is plotted for primary tumors (○) and nodal disease (●), as well.](image)

![Fig 3. Positive predictive value (PPV) of (18F)fluorodeoxyglucose positron emission tomography (FDG-PET)/computed tomography (CT) by risk group. The PPV of FDG-PET/CT for primary and nodal disease combined is shown for patients with oropharyngeal (OP) and non-OP primaries, tobacco users and nonusers, and tumors negative or positive for human papillomavirus (HPV). The number of patients in each group is displayed below each data point.](image)
high risk (n = 31). The rate of response to radiotherapy was greater in the low-risk group compared with the high-risk group (95.1% vs 71.0%, respectively).

ROC curve analysis demonstrated that the accuracy of FDG-PET/CT was greater for the high-risk group compared with the low-risk group (Fig 4). This difference was most pronounced for the detection of residual nodal disease, where the area under the ROC curves was 50% greater for high-risk versus low-risk patients.

Finally, we compared the accuracy of both imaging modalities across the two risk groups (Table 2). The most pronounced benefit achieved by using FDG-PET/CT comes in detecting residual nodal disease for high-risk patients, where the positive predictive value of FDG-PET/CT is twice that of CT alone (75% vs 37.5%, respectively). FDG-PET/CT also increases the specificity of response assessment compared with CT alone for high-risk patients, at the cost of slightly lower sensitivity and negative predictive value. For the low-risk group, FDG-PET/CT was inferior to CT alone for all metrics of response assessment accuracy.

**DISCUSSION**

Several retrospective series have investigated the utility of FDG-PET/CT imaging in HNSCC radiation response assessment.13-20,22 These series are confounded by limitations inherent to their retrospective design and by small sample size. A number of small prospective studies have been published using FDG-PET alone, without coregistered CT imaging, yielding inconsistent results.19,26,27 Two of these studies directly correlated imaging results with neck dissection findings; these demonstrated widely ranging staging accuracy (negative predictive values of 14% and 91.7%).19,27

To our knowledge, the current study is the first to prospectively evaluate combined FDG-PET/CT for this indication. Our data confirm that FDG-PET/CT imaging provides little value over conventional CT imaging for assessment of radiation response in unselected patients with locally advanced head and neck cancer. Our analytic approach was to weight the cost of sensitivity and specificity equally to generate SUVmax thresholds with maximal overall accuracy. Doing so, FDG-PET/CT outperformed CT only in terms of specificity and positive predictive value. These characteristics are best suited for minimizing risk of unnecessary neck dissections. Most patients (72.7%) who underwent consolidative neck surgery on this protocol did so without clinical benefit, and there was no evidence that FDG-PET/CT findings influenced the rate or type of neck dissection performed (selective or complete). The increased specificity of FDG-PET/CT would seem suited to address this problem. Extrapolating from the data presented earlier, FDG-PET/CT could reduce the rate of false positives by approximately one third, eliminating false-positive imaging results for 12 patients in every 100 patients treated. Whether this benefit
outweighs a 67% relative increased risk of false-negatives findings with FDG-PET/CT is worthy of debate. Our bias is to accept the risk of unnecessary consolidative dissections to avoid missing residual nodal disease. Given these data, then, we would not recommend routine use of FDG-PET/CT for radiation response assessment in unselected HNSCC patients.

Despite such findings, subset analysis revealed that FDG-PET/CT can provide benefit when applied in an individualized, risk-adjusted fashion. We initially surmised that selective use of FDG-PET/CT response assessment in patients at highest risk for recurrence would improve its diagnostic yield. Consistent with that premise, we demonstrated that FDG-PET/CT improved on standard CT response assessment for patients with HPV-negative tumors, nonoropharyngeal primaries, and histories of alcohol and tobacco use—patients who were at elevated risk for treatment failure. For these high-risk patients, nodal PET positivity after radiotherapy likely warrants surgical salvage (positive predictive value = 75%). A negative study should not necessarily preclude dissection of a neck where salvage is otherwise indicated because one in 20 of these patients would be in need of further therapy (negative predictive value = 95%).

The initial impetus to investigate HPV status in our patient population came from the observation that a history of heavy alcohol and tobacco use correlated with poor response to therapy on this trial. Tobacco use has been formally associated with tumor population came from the observation that a history of heavy smoke but has been formally associated with tumor HPV negativity in a recent population-based study,28 and tumor HPV negativity correlates strongly with poor response to radiotherapy29 and shorter overall survival.30 We must caution that less than half of patients on this study had sufficient tissue available for analysis. This is a consequence of patients commonly presenting to our institution after a tissue diagnosis had been made at an outside facility. Nevertheless, our expanded subset analysis showed that other factors, namely nonoropharyngeal primary tumors and a history of significant tobacco use, also correlated with increased accuracy of FDG-PET/CT–based response assessment. We feel that the ability to triage patients based on primary tumor site and tobacco use is of great clinical relevance, given that HPV genotyping is not available universally.

Several limitations of this study merit attention. First, because this study was performed at a high-volume cancer center, the results may not be applicable to all treatment settings. The referral pattern at our institution is potentially different from other settings, where the proportion of high-risk patients and radiotherapy nonresponders may be greater. Our results suggest that FDG-PET/CT would be more accurate in such a scenario and may, therefore, deserve broader application in that setting. Second, this study was not designed to evaluate the impact of post-treatment interval on the accuracy of FDG-PET/CT, so this remains an open question.

The approach at our institution is to assess response 8 to 9 weeks after radiotherapy. This permits time for follow-up examinations by all providers, formal reading of imaging studies, and arrangement of necessary surgery within the time window preferred by our surgeons to minimize operative morbidity in the postradiation setting. Several groups have suggested that the accuracy of FDG-PET/CT scanning may increase with longer (10 to 12 weeks) postradiotherapy intervals.15,18,26,31 Other authors have described 8 weeks after radiotherapy as an appropriate cutoff interval to maximize accuracy.15,32 When our own data were stratified by post-treatment interval, we observed no clear benefit to delaying imaging beyond 8 weeks (Appendix Figs A2 and A3, and Table A3, online only).

In conclusion, we prospectively demonstrate that FDG-PET/CT provides little value over CT alone in radiation response assessment for unselected patients with locally advanced HNSCC. Nonetheless, our data also suggest that FDG-PET/CT may improve assessment of treatment response in high-risk patients, such as those with HPV-negative disease. The currently enrolling Radiation Therapy Oncology Group 0522 phase III trial has incorporated a formal study of FDG-PET/CT imaging for treatment response assessment that closely mirrors the design of this current study. Confirmation of our results in the cooperative group setting would provide critical impetus to incorporate risk stratification strategies into FDG-PET/CT assessment of radiotherapy response in locally advanced head and neck cancer.

Although all authors completed the disclosure declaration, the following author(s) indicated a financial or other interest that is relevant to the subject matter under consideration in this article. Certain relationships marked with a “U” are those for which no compensation was received; those relationships marked with a “C” were compensated. For a detailed description of the disclosure categories, or for more information about ASCO’s conflict of interest policy, please refer to the Author Disclosure Declaration and the Disclosures of Potential Conflicts of Interest section in Information for Contributors.

**Employment or Leadership Position: None Consultant or Advisory Role: Homer A. Macapinlac, GE Healthcare (C), Radiology Corporation of America (C) Stock Ownership: None Honoraria: Homer A. Macapinlac, GE Healthcare, Siemens, Radiology Corporation of America Research Funding: Homer A. Macapinlac, GE Healthcare Expert Testimony: None Other Remuneration: None**

**AUTHORS’ DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST**

Conception and design: Homer A. Macapinlac, K. Kian Ang, David L. Schwartz

Financial support: K. Kian Ang, Scott M. Lippman, David L. Schwartz

Administrative support: David L. Schwartz


Collection and assembly of data: Benjamin J. Moeller, Vishal Rana, Blake A. Cannon, Michelle D. Williams, Erich M. Sturgis, Lawrence E. Ginsberg, David L. Schwartz


Manuscript writing: Benjamin J. Moeller, Erich M. Sturgis, Homer A. Macapinlac, J. Jack Lee, David L. Schwartz


**AUTHOR CONTRIBUTIONS**
REFERENCES


