IMPORTANCE No guidelines at present describe when fludeoxyglucose F 18–labeled positron emission tomography and computed tomography (FDG PET-CT) should be used in the initial posttreatment period for evaluation of oropharyngeal squamous cell carcinoma treatment outcome and recurrence.

OBJECTIVE To compare accuracies of the initial posttreatment PET-CT between primary treatment groups and to define indicators of false-positive findings.

DESIGN, SETTING, AND PARTICIPANTS This retrospective cohort study identified adults with a new diagnosis of oropharyngeal squamous cell carcinoma who received treatment with curative intent from October 1, 2006, through November 30, 2016, using the Alberta Cancer Registry (n = 380). Patients who underwent PET-CT within 1 year of treatment completion were included (n = 190). Of these, 103 patients (54.2%) had PET-CT findings positive for residual or recurrent disease, and 61 (32.1%) had false-positive findings. Among the 61 patients, 42 (68.9%) had received chemoradiotherapy (CRT) and 19 (31.1%) had primary surgery. Forty-two patients had true-positive findings, indicating a prevalence rate of disease of 22.1%. Data were analyzed from July through October 2017.

EXPOSURES One of 2 primary treatment modalities (surgery with or without adjuvant therapy vs CRT). All patients had posttreatment FDG PET-CT.

MAIN OUTCOMES AND MEASURES Primary outcome measures included the diagnostic odds ratio, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of PET-CT for detecting residual and/or recurrent disease. A multivariate analysis determined indicators of false-positive findings. Discriminative ability was assessed using receiver operating characteristic curve analysis of maximum standardized uptake value (SUVmax) metabolic data.

RESULTS Of the 190 participants, 77.9% were men, with a mean (SD) age at diagnosis of 58.5 (8.5) years. The diagnostic odds ratio was 19.3 (95% CI, 5.7-65.1); pooled sensitivity, 93.3% (95% CI, 80.7%-98.3%); and pooled specificity, 57.9% (95% CI, 49.4%-66.0%). The PPV of detecting disease was 54.7% (95% CI, 38.8%-69.8%) for primary surgery and 31.1% (95% CI, 20.2%-44.4%) for CRT. The NPV was 100% (95% CI, 94.7%-100%) for primary surgery and 96.6% (95% CI, 89.5%-99.1%) for CRT. Multivariate analysis identified treatment type, p16 disease, and smoking status as indicative of false-positive findings. In the receiver operating characteristic curve analysis for primary tumors, the optimal cutoff SUVmax for indicating true- vs false-positive results was 5.1 for surgically treated patients (area under the curve, 0.729; 95% CI, 0.570-0.888) and 5.3 for patients treated with CRT (area under the curve, 0.844; 95% CI, 0.700-0.989).

CONCLUSIONS AND RELEVANCE The results indicate a higher specificity for FDG PET-CT for initial posttreatment surveillance imaging among patients treated with primary surgery compared with nonsurgical management. Both sets of patients with posttreatment FDG PET-CT findings with an SUVmax greater than 5.0 should undergo close evaluation for possible residual or recurrent disease.
The rate of locoregional or distant recurrence remains high in head and neck squamous cell carcinoma (HNSCC). As many as 90% of these recurrences occur within the first 2 years after treatment, emphasizing the need for an accurate posttreatment surveillance program.\(^1\)\(^2\) The type of program, however, including its interval and duration in the initial posttreatment period, remains undefined.

Owing to its difficult anatomical location and treatment-associated changes to normal anatomy and tissues, physical examination alone has been proved to be insufficient as a surveillance method for oropharyngeal squamous cell carcinoma (OPSCC).\(^2\) Thus, clinical follow-up has been supplemented with radiologic imaging to increase the diagnostic accuracy of detecting true residual and/or recurrent disease.

Fludeoxyglucose F 18-labeled positron emission tomography and computed tomography (FDG PET-CT) have emerged as an advanced imaging modality in the assessment of HNSCC. PET-CT offers functional and anatomical information that has improved diagnostic accuracy in the pretreatment and posttreatment settings. A meta-analysis of 27 reports on the utility of PET scans for the posttreatment follow-up of HNSCC showed an overall pooled sensitivity and pooled specificity of 86.2% and 82.3%, respectively, for the detection of residual or recurrent disease at the primary site.\(^3\) The negative predictive values (NPVs) were 95% for the primary site and 96% for neck disease, whereas the positive predictive values (PPVs) were 75% for the primary site and 49% for the neck. Similar values were obtained in a study of 80 patients with head and neck cancer treated with radiotherapy.\(^4\) With a prevalence rate of disease of 23.8%, a negative (57 of 80 patients) vs positive (22 of 80 patients) PET-CT finding within 6 months of treatment completion offered a significant prognostic value (3-year overall survival, 100% vs 32%; \(P = .01\)).\(^4\)

The \(^{18}\)F-FDG tracer is a radiopharmaceutical analogue of glucose metabolism that reflects cellular activity and proliferation.\(^5\) Metabolic variables allow for a noninvasive and quantitative assessment of tumor burden by analyzing the 3-dimensional distribution of activity based on the photons emitted by the labeled tracer.\(^6\) The standardized uptake value (SUV) is the most widely used method for quantifying \(^{18}\)F-FDG uptake. The maximum SUV (SUVmax) of primary HNSCCs has been shown to be significantly associated with recurrence and patient survival.\(^7\) As a unique marker of tumor burden, the SUVmax allows for risk stratification of patients, lending itself as a possible contributor in clinical decision-making algorithms regarding suspected recurrences in patients with treated HNSCC.

Despite this information, the interpretation of \(^{18}\)F-FDG activity is skewed by physiological uptake in normal organs and benign tumors.\(^8\) Moreover, the inflammatory processes that occur in treated patients after surgery or in those receiving systemic therapy contribute to the false-positive rate seen in posttreatment PET-CT assessment.\(^8\)

The goal of a head and neck cancer surveillance program should be to achieve earlier detection of recurrence, with increased rates of successful salvage treatment and improved survival.\(^9\) Despite the reported high sensitivities of PET-CT imaging for detecting residual and/or recurrent disease, a subset of patients still undergo invasive salvage therapies secondary to false-positive findings. In addition, previous studies have not compared the diagnostic accuracy of PET-CT imaging between primary treatment modalities. The primary objective of our study was to measure the accuracy of initial posttreatment PET-CT between primary treatment groups (surgery with or without adjuvant therapy vs chemoradiotherapy [CRT]) for detecting residual and/or recurrent disease in patients with OPSCC. We also sought to outline the indicators of false-positive imaging findings in patients with oropharyngeal cancer treated with nonsurgical therapies.

### Methods

**Patient Population**

A retrospective cohort study was performed of adult patients (>17 years) who underwent treatment for biopsy-proved OPSCC from October 1, 2006, to November 30, 2016. Patients were identified using a prospectively collected database (Alberta Cancer Registry) that, by legal mandate, incorporates information on all patients treated at any tertiary cancer treatment facility in the province of Alberta, Canada. The patients underwent staging according to the 7th edition of the American Joint Committee on Cancer criteria. The University of Alberta Research Ethics Board approved this study and waived the need for informed consent for use of publicly available data. All patients were treated with primary surgery with or without adjuvant treatment or primary CRT. Only those who were treated with curative intent and underwent PET-CT within 1 year of treatment completion were recruited. Exclusion criteria were (1) distant metastatic disease; (2) all other major head and neck cancers; (3) loss to follow-up; and (4) diagnosis at our institution with treatment elsewhere.

All treatment followed the 2015 National Comprehensive Cancer Network Guidelines for management of oropharyngeal cancer.\(^10\) Surgical treatment consisted of transoral, transoral robotic, transmandibular, and/or transcervical resections of the primary tumor site(s). When required, locoregional and/or free flap removal was performed.

### Key Points

**Question** Can the specificity of positive positron emission tomographic and computed tomographic findings in the setting of surveillance of patients treated for oropharyngeal squamous cell carcinoma be improved with different modalities?

**Findings** In this retrospective cohort study of patients treated for oropharyngeal head and neck cancer, specificity was higher for the initial posttreatment positron emission tomography and computed tomography in patients who underwent primary surgery compared with nonsurgical management. Treatment type, p16 status, and smoking status were indicative of false-positive results, and a maximum standardized uptake value was double the current standard for differentiating between true- and false-positive findings.

**Meaning** According to results of this study, caution should be used in interpreting initial posttreatment positron emission tomographic and computed tomographic findings in patients with oropharyngeal cancer treated with nonsurgical therapies.
FDG PET-CT in the Detection of Oropharyngeal Cancer Recurrence

Original Investigation Research

reconstruction was used to close larger oropharyngeal defects. The types of free flap reconstruction included radial forearm or anterolateral thigh flaps. The primary tumors were excised with margins of 1 cm or greater (peripheral and deep margins). Bilateral selective neck dissections (levels I-V) were performed in patients with clinically positive nodal disease, with levels IV and V explored if indicated. Currently, our institution follows the treatment protocol with staging based on the 7th edition of the American Joint Committee on Cancer staging guidelines. All patients with early-stage OPSCC (stages I-II) regardless of p16 or smoking status are offered (where appropriate) single-modality treatment consisting of radiotherapy or surgery. Patients with advanced-stage OPSCC (overall stage III or IV) are offered combined-modality treatment with a preferred pathway based on p16 and smoking status (positive smoking status defined as >20 pack-years for patients positive for p16 and >10 pack-years for patients negative for p16). Nonsmokers who were positive for p16 were offered CRT; smokers who were positive or negative for p16 were offered primary surgery with adjuvant radiotherapy with or without chemotherapy (based on high-risk features on final histopathologic evaluation); and nonsmokers negative for p16 were offered CRT or surgery with adjuvant radiotherapy and/or chemotherapy (based on high-risk features on final histopathologic evaluation).11,12

Postoperative CRT was performed in the presence of pathologic risk factors. The indications for radiotherapy included the following: T3-T4 N2-N3 disease, positive margins, perineural invasion, lymphovascular invasion, and extracapsular spread. Concomitant CRT with cisplatin-based regimens was administered to patients presenting with extracapsular spread and positive margins. As per standard treatment practices of external intensity-modulated radiotherapy, the radiation field included the entire tumor bed area (with 1- to 2-cm margins) as well as the regional lymphatics at risk.

Imaging Studies and Evaluation

The PET-CT imaging was performed per our institutional head and neck protocol. The PET-CT protocols were obtained using two 16-section PET-CT scanner systems (Gemini TF; Phillips Healthcare). Patients fasted from more than 4 to 6 hours, and serum samples for baseline glucose measurements were collected. A minimum of 60 minutes was waited after intravenous administration of 5.18 MBq/kg of 18F-FDG tracer. We performed PET imaging using 2 separate acquisitions from skull vertex to midthigh. After the PET scan, patients were injected with intravenous contrast, and a helical CT scan was performed to scan patients from skull vertex to midthigh.

All FDG PET-CT images were interpreted by a board-certified nuclear medicine physician at the time of imaging. The scan reports were retrospectively read and the results were classified as true- or false-positive or true- or false-negative. A negative designation was given when no evidence of disease was found on the CT or the PET portion of the examination. Asymmetrical 18F-FDG uptake where recurrent or residual tumor could not be excluded was interpreted as a positive finding. Positive findings on scans were correlated with subsequent confirmation via histopathologic analysis, radiographic clearance, or clinical correlation (ie, no evidence of residual and/or recurrent disease within 1 year of PET-CT).

For SUVmax values, a joint reading of the head and neck CT and FDG PET scans was performed side by side to avoid the inclusion of areas with physiological FDG uptake within the regions of interest. The tumor boundaries were defined using a fixed SUV threshold of 2.5, measured in a volume of interest drawn around the primary tumor. The SUVmax values were automatically calculated by the software.

Demographic features, smoking history (±10 pack-years), p16 status, disease staging, primary treatment, PET-CT result, and treatment effect of PET-CT were obtained from the Alberta Cancer Registry and verified by retrospective review of patients’ medical records. Some p16 data were missing in the CRT cohort (because it was not the protocol to stain for p16 positivity, especially for those treated with primary CRT, until recently).

Statistical Analysis

Data were analyzed from July through October 2017. We calculated the pooled sensitivity, pooled specificity, PPV, NPV, and diagnostic odds ratio (OR). Multivariate analysis was conducted using the binary logistic regression (backward stepwise) method. The diagnostic performance of SUV variables in differentiating true- from false-positive PET-CT findings was analyzed using receiver operating characteristic curves and area under the curves with 95% CIs.

Results

Patient Cohort

Of the 380 patients with OPSCC, 190 had initial posttreatment PET-CT within 1 year of treatment completion (148 men [77.9%] and 42 women [22.1%]; mean [SD] age at diagnosis, 58.5 [8.5] years). Analysis of baseline demographic data revealed no significant differences between patients in the 2 primary treatment modality groups in age at diagnosis, sex, p16 status, or overall staging (Table 1). The tumor site included the oropharynx in 6 patients; tonsil, 96; base of the tongue, 85; and palate, 3. Initial tumor category was T1 in 44 patients; T2, 73; T3, 40; and T4, 33. One hundred three of 190 patients (54.2%) had PET-CT positive findings of residual and/or recurrent disease, whereas the remaining 87 patients (45.8%) had PET-CT findings of no residual and/or recurrent disease (Table 2). Sixty-one patients had false-positive findings (32.1%), of whom 42 (68.9%) had received primary CRT, and 19 (31.1%) received primary surgical interventions. Forty-two patients had true-positive findings, indicating a prevalence rate of 22.1%. Of the 61 patients with false-positive findings, 23 were cleared clinically; 14 underwent quadroscopic examination under anesthesia with biopsy; 6 underwent neck dissections; and 18 had follow-up radiologic imaging (a CT scan in a mean of 1.65 [0.96] months or a PET-CT scan in a mean of 3.70 [1.27] months).

The pooled sensitivity for detecting OPSCC was 93.3% (95% CI, 80.7%-98.3%); pooled specificity was 57.9% (95% CI, 49.4%-66.0%) (Table 3). The PPV for detecting residual and/or recurrent disease in patients undergoing surgical manage-
FDG PET–CT in the Detection of Oropharyngeal Cancer Recurrence

The purpose of obtaining PET-CT scans as a surveillance tool is to allow for the early detection of recurrent disease (in the primary site and the neck), to assess for a metachronous second primary tumor, and to rule out distant metastases. The current use of PET-CT for restaging a suspected recurrence continues to be chiefly driven by clinical symptoms. No specific guidelines in the literature define the indications or timing of the appropriateness of PET-CT in the initial posttreatment period. With respect to the radiologic surveillance and management of patients with posttreatment OPSCC, one must first clarify (1) the role of PET-CT in detecting residual or recurrent disease; (2) the criteria that are useful in interpreting the results of a positive initial posttreatment PET-CT finding; and (3) treatment options after a positive initial posttreatment PET-CT finding.

Table 1. Characteristics of All Patients With OPSCC Included in the Study

<table>
<thead>
<tr>
<th>Characteristic at Time of Initial Diagnosis</th>
<th>Treatment Modality</th>
<th>Effect Size (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Primary Surgery (n = 74)</td>
<td>Primary CRT (n = 116)</td>
</tr>
<tr>
<td>Age, mean (SD), y</td>
<td>58.4 (8.1)</td>
<td>58.6 (8.7)</td>
</tr>
<tr>
<td>Male, No. (%)</td>
<td>55 (74.3)</td>
<td>94 (81.0)</td>
</tr>
<tr>
<td>p16-Positive status, No. (%)</td>
<td>47 (63.5)</td>
<td>85 (73.3)</td>
</tr>
<tr>
<td>Overall stage, No. (%)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stage</th>
<th>Primary Surgery</th>
<th>Primary CRT</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>2 (2.7)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>1 (1.4)</td>
<td>3 (2.6)</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>8 (10.8)</td>
<td>16 (13.8)</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>63 (85.1)</td>
<td>97 (83.6)</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Posttreatment PET-CT and Disease Status

<table>
<thead>
<tr>
<th>PET-CT</th>
<th>Disease, No. (%)</th>
<th>All Patients, No. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Present Absent</td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>42 61 103</td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>3 84 87</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>45 145 190</td>
<td></td>
</tr>
</tbody>
</table>

The PPVs and NPVs are directly associated with the prevalence of the disease in the population. The PPV increases with an increase in the prevalence of disease, whereas the NPV decreases with an increase in the prevalence of disease. Our prevalence rate of disease is low (22.1%), which is consistent with our study findings of a low PPV and a high NPV.

The Role of PET-CT in Posttreatment Surveillance

Our study results are consistent with those of existing literature on the diagnostic variables of PET-CT for detecting residual and/or recurrent disease in patients with oropharyngeal cancer. Several small, single-institution prospective and retrospective series have demonstrated an NPV of 90% to 100% for the detection of persistent nodal metastasis after radiotherapy or CRT, with a variable PPV ranging from 30% to 60%. These values are notably higher values than the PPV of CT and/or magnetic resonance imaging and better than combined clinical examination and ultrasonography, cementing PET-CT as an optimal tool for surveillance imaging.

The PPVs and NPVs are directly associated with the prevalence of the disease in the population. The PPV increases with an increase in the prevalence of disease, whereas the NPV decreases with an increase in the prevalence of disease. Our prevalence rate of disease is low (22.1%), which is consistent with our study findings of a low PPV and a high NPV.

Criteria for Interpreting the Results of a Positive Initial Posttreatment PET-CT Finding

In this cohort study, we evaluated the accuracy of PET-CT in detecting residual or recurrent disease between primary treatment modalities in OPSCC. Our results suggest that PET-CT in patients treated with CRT has a lower PPV for detecting disease in the initial posttreatment period, yielding significantly more false-positive findings than in patients treated with primary surgery.

Furthermore, we found smoking and p16 status to be associated with posttreatment findings. In patients with a positive posttreatment scan finding, those with p16-negative disease or with a smoking history of at least 10 pack-years were significantly more likely to have false-positive findings.
more likely to have a true-positive finding of residual and/or recurrent disease. Time to PET-CT and diabetes were not correlated with false-positive findings.

To our knowledge, this study is the first of its kind to compare the diagnostic accuracy of PET-CT between treatment modalities. Anecdotally, we noticed a significant number of salvage surgical procedures with negative pathologic findings in patients treated with CRT, secondary to an initial false-positive PET-CT finding. Our study results are in keeping with this observation, demonstrating a lower PPV in this subset of patients than in those treated with primary surgery. The leading hypothesis to explain this phenomenon lies within the nature of the treatment itself. Surgical treatment involves tumor resection, which removes the hypermetabolic tumor, whereas systemic CRT results in cellular changes that involve significant inflammation. With morphologic imaging, differentiation between the presence of treatment-induced changes and residual or recurrent disease is difficult (eFigure in the Supplement).

Analysis of metabolic variables offers an opportunity to estimate individual tumor behavior, including the likelihood of aggressive disease, response to therapy, and patient outcomes. The current literature suggests an isocontour threshold SUVmax value of 2.5 for differentiating malignant vs benign cellular activity in the pretreatment setting. Our study suggests a cutoff value double the current standard for differentiating between true- and false-positive scan findings, regardless of treatment type. This finding bears implications for the clinical management of head and neck cancer and for prognostic stratification. Using specific metabolic variables can assist with a watch-and-wait approach guided by PET-CT, especially in those patients who may not be ideal surgical candidates, compared with planned salvage therapy. For example, in our study, 20 patients underwent an invasive surgical procedure, 6 of whom had neck dissections based on PET-CT findings. Using PET-CT for the guided approach avoids the additional morbidity of further surgery in irradiated head and neck tissue, which is a considerable contribution to the management algorithm in head and neck oncology.

In our study, we found that the predictive value of a post-treatment PET-CT was lower with human papillomavirus (HPV)-associated disease. Our results are consistent with previously published studies. Clark et al found a significant association between a higher nodal SUVmax value and p16-positive disease status. In addition, Moeller et al found patients with high-risk features (negative for HPV, history of tobacco use) to have a nodal PPV of 75% vs only 14% (secondary to marked 18F-FDG activity) for those with low-risk features (positive for HPV, primary tumor of the oropharynx, and no tobacco use). These findings have been postulated to represent a functionally intact and responding immune system with high glycolytic activity in HPV-positive disease.

Time to PET-CT has previously been shown to be indicative of false-positive results. Studies have demonstrated 12 weeks after treatment as an appropriate time to maximize accuracy and reduce false-positive and false-negative findings. Our study did not find a significant difference in false-positive findings with posttreatment time of imaging. However, the relatively smaller cohort size, especially after stratification into groups based on time to posttreatment PET-CT, places severe limits on the widespread generalizability of this finding.

However, we did find a significant difference when patients who received PET-CT scans at no more than 3 months were stratified by treatment type. More false-positive findings were detected in patients who received primary CRT compared with the primary surgery cohort. This finding is consistent with previously published literature on CRT. Wang et al evaluated nodal response and PPV of the 3-month posttreatment PET scan in patients with HPV-associated oropharyngeal cancer treated in a multi-institutional deintensification trial. The PPV of posttreatment scans ranged from 9% to 13%, leading the authors to recommend continued surveillance rather than surgical evaluation. This PPV range was comparatively lower than the PPV of 66.8% for the posttreatment PET scan obtained in patients treated with
primary surgery (with a median interval of 3.4 months). A balance must be drawn between maximizing accuracy and detecting residual disease earlier to facilitate timely salvage treatment. However, the overall recommendation remains that the initial PET-CT scan is ideally performed 12 weeks after treatment completion.

Treatment Options After a Positive Initial Posttreatment PET-CT Finding

The treatment of residual disease as indicated by a positive initial posttreatment PET-CT finding is currently determined by the treating clinician. Anatomical correlates and an isocountour threshold SUVmax value of 2.5 are used to estimate malignant activity. Values that are closer to this SUVmax limit are considered equivocal for recurrent or persistent disease, leading clinicians to choose less invasive means of evaluating patients (ie, clinical clearance through regular follow-up at a minimum of 3 months with a careful physical examination, or follow-up radiologic imaging with CT or PET-CT scanning). In contrast, PET-CT imaging indicating intense FDG uptake, with or without anatomic correlates, often leads to invasive strategies to determine the presence or absence of disease (ie, quadrosopic examination under anesthesia with biopsy, or salvage neck dissections). No established PET-CT guidelines exist in the literature that help clinicians choose between these treatment options. The decision-making algorithm would be facilitated considerably by the institution of posttreatment PET-CT variables or, as our study would indicate, an SUVmax value of 5.0.

Limitations

We acknowledge that our study has several limitations. First, this was a retrospective, single-center study; second, protocols for PET-CT scans of the head and neck vary by institution, with different radiologic softwares based on institutional preferences. Our study results, specifically with respect to metabolic variables as well as errors inherent in SUV definition, should be interpreted considering this heterogeneity.

Conclusions

To our knowledge, this study is the first to examine the role of the initial posttreatment PET-CT scan when comparing different treatment modalities (primary surgery vs primary organ preservation) and to demonstrate novel findings regarding PET-CT interpretation compared with the current literature. In surgically and nonsurgically treated cohorts, an SUVmax of greater than 5.0 was found to be the threshold cutoff value between true- and false-positive findings of persistent disease. Further exploration of these findings is necessary, including a multicenter cohort analysis if necessary, before widespread adaptation of an SUVmax value of 5.0 (regardless of treatment subtype) to guide surgical decision making. The findings described herein have the potential to significantly reduce unnecessary surgical interventions and treatments in an already challenging-to-treat patient population.
REFERENCES