**Head and neck squamous cell cancer (stages III and IV) induction chemotherapy assessment: Value of FDG volumetric imaging parameters**

Jielin Yu,1 Timothy Cooley,2 Minh-Tam Truong,3 Gustavo Mercier1 and Rathan M. Subramaniam1,3*

1Department of Radiology, 2Medical Oncology, 3Radiation Oncology, Boston University School of Medicine, Boston, Massachusetts, USA

**Abstract**

**Introduction:** To evaluate whether the change in the metabolic tumour volume (MTV) or total lesion glycolysis (TLG) of the primary tumour, before and after induction chemotherapy, predicts outcome for patients with advanced head and neck squamous cell cancer (SCC).

**Methods:** Twenty-eight patients with advanced (American Joint Committee on Cancer stage III and IV) head and neck SCC who underwent positron emission tomography (PET)/CT were included in this retrospective study. Primary tumour MTV and TLG were measured using gradient and fixed percentage threshold segmentations. Outcome endpoint was disease progression or mortality. Pearson correlation, Bland–Altman and receiver operator characteristic analysis were performed.

**Results:** The Pearson’s correlation coefficients between percentage changes (pre- and post-induction chemotherapy) from gradient MTV (MTVG) and the 38% SUVmax threshold MTV (MTV38) was 0.96 and between MTVG and the 50% threshold MTV (MTV50) was 0.95 ($P < 0.0001$). The corresponding Pearson $r$ between TLGG and TLG38 was 0.94 and between TLGG and TLG50 was 0.96 ($P < 0.0001$). The least bias was 1.89% (standard deviation = 25.30%) between the percentage changes of MTVG and MTV50. The areas under the curve for predicting progression or mortality were 0.76 ($P = 0.03$) for MTVG and 0.82 for TLGG ($P = 0.009$). Optimum cut points of a 42% reduction in MTVG and a 55% reduction in the TLGG predict event-free survival with a sensitivity of 62.5% and a specificity of 90% and a hazards ratio of 6.25.

**Conclusion:** A reduction in primary tumour MTV of at least 42% or in TLG of at least 55% after induction chemotherapy may predict event-free survival in patients with advanced head and neck SCC.

**Key words:** head and neck; PET/CT.

**Introduction**

Induction chemotherapy is an emerging treatment paradigm for treating inoperable locally advanced squamous cell carcinomas of the head and neck.\(^1\) Induction chemotherapy followed by chemoradiotherapy is a form of treatment intensification and is evolving. This may contribute to a better survival rate compared with chemoradiotherapy alone.\(^2\) Monitoring the response to induction chemotherapy provides critical information that allows physicians to triage to the best post-induction treatment.

Fluorodeoxyglucose positron emission tomography/computed tomography (FDG PET/CT) has proven to be clinically useful as a diagnostic, staging and therapy planning tool.\(^3\)–\(^7\) The role of FDG PET/CT in determining the outcome of induction chemotherapy has been explored. Previous studies indicate that PET/CT shows a significant correlation with pathological response following induction and concurrent chemoradiotherapy (CCRT).\(^8\) It has been shown to have the same efficacy as endoscopy with biopsy for determining the tumour volume after induction therapy.\(^9\)
The purpose of this exploratory study is to evaluate whether the change in the metabolic tumour volume (MTV) and total lesion glycolysis (TLG) of the primary tumour, before and after induction chemotherapy, predicts outcome of patients with locally advanced head and neck cancer.

Methodology

Patients and study design

This is a single-institution retrospective study and was approved by the Institutional Review Board. Informed consent requirements were waived for this study. All 28 patients, with biopsy-proven head and neck squamous cell carcinoma who underwent a FDG PET/CT at baseline and post-induction chemotherapy from January 2010 to April 2011 in our institution, were included in the study. All the patients were treated with induction chemotherapy followed by CCRT. The induction chemotherapy regimen consisted of three cycles of Taxotere (docetaxol), cisplatin and 5-fluorouracil (TPF), with minor adjustments for toxicity and other complications. Post-induction therapy consisted of 6 to 7 weeks of CCRT, with 69.96 Gy over 33 fractions.

PET/CT

All PET/CT studies were performed on a GE Discovery STE 16 (General Electric, Milwaukee, WI, USA) PET/CT scanner according to the institutional standard clinical protocol. For all patients, a dedicated head and neck protocol was instituted. Patients were scanned from skull base to aortic arch with the arms down and then clavicle to mid thigh with the arms up. The average patient blood glucose level was 103 mg/dl (standard deviation (SD) 21 mg/dl). Patients were injected with an average of 11.3 mCi (SD 0.6 mCi) or 418.1 M bq (SD 22.2 M bq) of 18F-FDG and incubated for an average period of 65 min (SD 7 min).

The dedicated head and neck imaging protocol consisted of two-dimensional PET scans obtained from the skull base to the arch of the aorta with a 30-cm field of view and 128 x 128 matrix. The emission scan lasted for 5 min per bed position. The remainder of the body (down to the mid thighs) was imaged using a weighted-basis emission scan time per bed position. PET slice thickness was 3.27 mm. Helical (16 detector) CT images were obtained with a matrix of 512 x 512. Beam collimation was 10 mm with a pitch of 0.984. Table speed was 9.84 mm/rotation, and the slice thickness was 0.625 mm. KV of 120 and mAs of 440 were used. CT images were reconstructed using a slice thickness of 3.75 mm every 3.75 mm. In addition, CT images were reconstructed using a slice thickness of 1.25 mm every 1.25 mm in soft tissue and a bone algorithm to generate a diagnostic level CT of the neck for review. PET/CT studies were performed at baseline before the start of induction chemotherapy and 2 to 3 weeks after the completion of induction chemotherapy.

Image analysis

All PET/CT studies were retrieved from the electronic archival system and reviewed on a MIMvista workstation (software version 4.1, MIM Software Inc., Cleveland, OH, USA) by a board certified radiologist with nuclear radiology and neuroradiology faculty member with 3 years experience as a faculty and head and neck PET/CT imaging. PET, CT and fused PET/CT images were reviewed in axial, coronal and sagittal planes. For the purposes of this study, the relevant imaging biomarker measurements were MTV and TLG from PET. MTV was defined as the tumour volume with FDG uptake segmented by a gradient-based method and fixed threshold methods at 38% and 50% of SUVmax. The TLG was defined as (MTV) x (SUVmean). The commercially available MIMvista software analysis suite (MIM Software Inc.) includes a contouring suite for radiation therapy planning and a PET/CT fusion suite. The edges of the primary tumour were automatically calculated and outlined in both segmentation methods. Once the primary tumour was segmented, MTV and TLG were semiautomatically calculated by the MIMvista software. Because there is greater intrareader variability measuring the FDG volumetric parameters (MTV and TLG) for smaller volume tumour, we did not segment the nodal metastasis. In addition, previous studies have demonstrated primary tumour volume than nodal volume as predictors of outcome. For these reasons, we only segmented the primary tumours.

Segmentation methods

(a) Gradient segmentation method

Gradient segmentation of tumour volume identifies the tumour on the basis of a change in count level at the tumour border. Complex segmentation methods have been proposed including denoising, deblurring, gradient estimation and watershed transformation. The gradient segmentation method used in MIMvista (version 4.1) has been previously described, and is simple and easy to use. It calculates spatial derivatives along the tumour radius and then defines the tumour edge on the basis of derivative levels and continuity of the tumour edge. The software relies on an operator-defined starting point near the centre of the lesion. As the operator drags the cursor from the centre of the lesion, six axes extend out, providing visual feedback for the starting point of gradient segmentation. Spatial gradients are calculated along each axis interactively, and the length of an axis is restricted when a large spatial gradient is detected. The six axes define an ellipsoid that is then used as an initial
bounding region for gradient detection. The reader can add regions until visually satisfied that the entire primary tumour is included in the contour.

**(b) Fixed percent threshold segmentation method**

The fixed SUV<sub>max</sub> threshold contouring method relies on including all voxels that are greater than a defined percent of the maximum voxel within an operator-defined sphere (in this study 38% and 50% of SUV<sub>max</sub>). We used these fixed thresholds as have been used in previous studies. Cross-sectional circles are displayed in all three projections (axial, sagittal and coronal) to ensure three-dimensional coverage of the primary tumour.

**Outcome endpoints**

The primary outcome measure is to establish whether the change in exploratory imaging markers, MTV and TLG, before and after induction chemotherapy, predicts overall survival and progression-free survival (PFS). Overall survival is defined as the time from initiation of therapy to death or to the most recent follow-up. Progression-free survival is defined as the time from initiation of therapy to the first documented progression at the primary site, at regional nodes, or at distant metastatic sites. Death from the primary cancer within a documented site of recurrence or progression or death from an unknown cause is considered death from local regional disease. Electronic medical records, imaging records and office visits at our institution were used to establish the overall survival and PFS. Outcome measures of the study were PFS or disease progression/mortality. The median follow-up time was 12.5 months.

**Statistical methods**

We present our summary statistics as the mean ±SD or range for continuous variables, or frequency and percentage for categorical variables. We used the Pearson correlation coefficient to establish the relationship between different segmentation methods, and we used Bland–Altman analysis between the two best-correlated segmentation methods to establish the reliability of the methods for measuring changes in MTV and TLG. The baseline MTV or TLG was used as the denominator for the calculation of percentage reduction. Between group analysis was performed using the Mann–Whitney U-test. Receiver operating characteristic (ROC) analysis was used to determine area under the curve (AUC) to estimate the accuracy and predictive ability of MTV and TLG. The optimum cut points were established to predict event-free survival. We used the Prism 5 (GraphPad Software Inc., San Diego, CA, USA) and SPSS 19 (SPSS Inc., Chicago, IL, USA) statistical packages for all analyses, and all hypothesis tests are two-sided with a significance level of 0.05.

**Results**

**Patients**

The mean age of the patients was 59 years (range, 29–82 years). The primary tumour sites were oral cavity and oropharynx (n = 14), larynx (n = 5), hypopharynx (n = 5) and other (n = 4). The distribution of patients by the American Joint Committee on Cancer (6th edition) stage of the cancers included stages III (n = 3, 10.7%), IVA (n = 20, 71.4%), IVB (n = 4, 14.3%) and IVC (n = 1, 3.6%). The mean follow-up period for the patients was 11 months (range, 2–26 months). Patient characteristics and therapy details are summarized in Table 1.

**Volumetric and metabolic changes and segmentation methods**

The Pearson correlation coefficient between the percentage changes (before and after induction chemotherapy) in gradient MTV (MTVG) and 38% SUV<sub>max</sub> threshold MTV (MTV<sub>38</sub>) was 0.96 (P < 0.0001). The Pearson r between MTVG and 50% SUV<sub>max</sub> threshold MTV (MTV<sub>50</sub>) was 0.95 (P < 0.0001). The corresponding Pearson r between TLGG and TLG<sub>38</sub> was 0.94 (P < 0.0001) and 0.96 (P < 0.0001) between TLGG and TLG<sub>50</sub>.

The Bland–Altman analysis showed a bias of 4.48% (SD = 26.13%) between the percentage changes of MTVG and MTV<sub>38</sub>. The corresponding bias was 1.89% (SD = 25.30%) between the percentage changes of MTVG and MTV<sub>50</sub>. The corresponding biases for TLG were 6.51% (SD = 28.96%) and 5.63% (SD = 32.92%).

<table>
<thead>
<tr>
<th>Table 1. Patient characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Characteristic</td>
</tr>
<tr>
<td>Age, mean (range)</td>
</tr>
<tr>
<td>Sex</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td>Primary tumour site</td>
</tr>
<tr>
<td>Oral cavity/oropharynx</td>
</tr>
<tr>
<td>Larynx</td>
</tr>
<tr>
<td>Hypopharynx</td>
</tr>
<tr>
<td>Other sites</td>
</tr>
<tr>
<td>AJCC stage</td>
</tr>
<tr>
<td>III</td>
</tr>
<tr>
<td>IV</td>
</tr>
<tr>
<td>Completed three cycles of induction chemotherapy</td>
</tr>
<tr>
<td>Yes</td>
</tr>
<tr>
<td>No</td>
</tr>
<tr>
<td>Follow-up time, mean (range)</td>
</tr>
</tbody>
</table>
SUV\textsubscript{max} changes and outcomes
The median SUV\textsubscript{max} changes for patients who had no events and for patients who had progressed or died were 100\% (IQR 53.5\% to 100\%) and 51.88\% (IQR 19.07\% to 93.88\%), respectively (Mann–Whitney U-test \(P = 0.09\)). In the ROC analysis, the AUC for SUV\textsubscript{max} was 0.70 (95\% CI 0.47–0.92; \(P = 0.10\)).

MTV changes and outcomes
The median MTV\textsubscript{G} changes for patients who had no events and for patients who had progressed or died were 100\% (IQR 88.6\% to 100\%) and 36.75\% (IQR -55.19\% to 89.13\%), respectively (Mann–Whitney U-test \(P = 0.02\)) (Figs 1,2). The corresponding values using MTV\textsubscript{38} were 100\% (IQR 84.1\% to 100\%) and 19.05\% (IQR -44.1\% to 81.62\%); and for MTV\textsubscript{50} 100\% (IQR 84.85\% to 100\%) and 24.04\% (IQR -17.25\% to 77.78\%), respectively (Mann–Whitney U-test \(P = 0.02\)) (Figs 1,2). There was a significant difference between the two groups for both MTV\textsubscript{38} (\(P = 0.02\)) and MTV\textsubscript{50} (\(P = 0.02\)).

In the ROC analysis, the AUC for MTV\textsubscript{G} was 0.76 (95\% CI 0.55–0.97; \(P = 0.03\)) (Fig. 3). AUCs for MTV\textsubscript{38} and MTV\textsubscript{50} were 0.77 (95\% CI 0.56–0.98; \(P = 0.03\)) and 0.76 (95\% CI 0.55–0.98; \(P = 0.03\)). An optimum cut point of 42\% reduction in the MTV\textsubscript{G} predicts event-free survival with a sensitivity of 67\% and a specificity of 90\% and a hazards ratio (HR) of 6.3. The corresponding cut point for MTV\textsubscript{50} was a 53\% reduction with the same sensitivity, specificity and HR.

TLG changes and outcome
The median TLG\textsubscript{G} changes for patients who had no events and for patients who had progressed or died were 100\% (IQR 88.6\% to 100\%) and 43.16\% (IQR -55.47\% to 88.96\%), respectively (Mann–Whitney U-test \(P = 0.007\)) (Fig. 4). The corresponding values using TLG\textsubscript{38} were 100\% (IQR 94.34\% to 100\%) and 43.16\% (IQR -55.47\% to 88.96\%) (\(P = 0.007\)); and for TLG\textsubscript{50} 99.02\% (IQR 40.41\% to 100\%) and 41.96\% (IQR -21.98\% to 84.68\%) (\(P = 0.006\)).

In the ROC analysis, AUC for TLG\textsubscript{G} was 0.84 (95\% CI 0.66–1.00; \(P = 0.006\)) and 0.84 (95\% CI 0.65–1.00; \(P = 0.006\)). An optimum cut point of 55\% reduction in the TLG\textsubscript{G} predicts event-free survival with a sensitivity of 62.5\% and a specificity of 90\% and a HR of 6.25. The corresponding cut point for TLG\textsubscript{50} was 59\% reduction with the same sensitivity, specificity and HR.

Discussion
The aim of this exploratory study was to determine whether metabolic volumetric analysis of tumours before and after induction chemotherapy can predict the event-free survival of head and neck cancer patients. The results suggest that there is a significant difference in the changes of both MTV and TLG between patients who had progressed or died and those who had an event-free survival. Using ROC analysis, a 42\% reduction in MTV\textsubscript{G} or 55\% reduction in TLG\textsubscript{G} were determined to be optimal for predicting event-free survival. The AUCs for MTV\textsubscript{50}, TLG\textsubscript{G}, and TLG\textsubscript{50} were also determined and found to be similar to MTV\textsubscript{G}. In addition, the changes in percentage of metabolic volume segmented by gradient and by the fixed percentage threshold methods are highly correlated.

Several parameters have been utilized for response assessments. One conventional method is based on changes in the longest measured diameter of the tumour. The maximum diameter approach suffers from a relatively small degree of interobserver reliability.\(^\text{16,17}\) Additionally, the tumour diameter is only a unidimensional anatomical parameter and provides no metabolic or functional information about the tumour biology. A more functional metabolic parameter is the maximum SUV (SUV\textsubscript{max}), the pixel with the highest intensity in the tumour region. Compared with tumour diameter, SUV\textsubscript{max} has significantly less interobserver and intraobserver variability.\(^\text{17,18}\) However, several factors can affect the SUV\textsubscript{max} and may under- or overestimate the degree to which it depends on tumour size and the tumour-to-background ratio.\(^\text{19}\) The utility of SUV\textsubscript{max} as a prognostic factor has been explored with mixed results. With non-small-cell lung cancers and head and neck non-squamous cell carcinomas, using SUV\textsubscript{max} values has been
shown to have predictive value.\textsuperscript{20,21} Other studies, however, have indicated that SUV measurements are not significant or reliable prognostic factors.\textsuperscript{22,23}

In light of the shortcomings of tumour diameter and SUV\textsubscript{max}, a new method of assessing prognosis based on volumetric analysis has been drawing increased attention. In the case of small-cell lung cancers, the MTV has been demonstrated to be a significant predictor of patient outcome.\textsuperscript{24} MTV has also been shown to be an important independent predictor of disease-free survival in the case of oesophageal cancer as well as head and neck cancers.\textsuperscript{25,26} It has been suggested to add value to

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure2.png}
\caption{(a and d) Axial positron emission tomography (PET), (b and e) axial PET/CT and (c and f) coronal PET/CT of a 78-year-old woman with T3N2cM0 stage IVA squamous cell cancer of the oropharynx. Pre-induction chemotherapy primary tumour MTV (gradient segmentation) is 8.3 mL. Induction therapy consisted of three cycles of Taxotere (docetaxol), cisplatin and 5-fluorouracil. The post-induction primary tumour metabolic tumour volume is 6.6 mL (20% reduction). Patient had progressed 4 months after the completion of treatment and died 21 months after treatment.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure3.png}
\caption{Metabolic tumour volume (MTV) and total lesion glycolysis (TLG) (gradient segmentation) percentage change (pre- and post-therapy) receiver operating characteristic analysis for event-free survival. AUC, area under the curve.}
\end{figure}
staging in predicting prognosis of patients with oral and oral cavity squamous cell cancers. Our results further demonstrate that MTV and TLG can potentially be used for induction therapy assessment in head and neck squamous cell cancer. In contrary to a prior study, we used the same threshold cut points, 38% and 50% of the measured SUVmax in the baseline and then in the post-induction images, as the SUVmax of the primary tumour are likely to change with induction chemotherapy. Our study showed there is significant reduction in the threshold MTV and TLG, after induction chemotherapy, in those who had an event and those who did not. This may be due to lack of radiation inflammation in our study. We did not use the baseline SUVmax for threshold cut point in the post-induction images, as used by a prior study, as SUVmax changes with induction therapy.

Our results need to be interpreted in the context of our study design. This is an exploratory study with a small number of patients, and the MTV and TLG were segmented using single vendor-provided, FDA-approved, commercially available software. Segmentation of small residual tumours may not be accurate, especially when the tumour volume is less than 10 mL. We did not perform the analysis for nodal disease, as small volume disease have more variability in measurements but acknowledge nodal disease is an important prognostic factor. Though our study only included patients with advanced head and neck cancer (stage III and stage IV) patients, we did not perform Human Papilloma Virus (HPV) status routinely in all head and neck patients at the time of this study. Hence, HPV status is not included in the analysis, which may influence the outcome of some of the patients included in the study. Due to the small sample size, we included both oral cavity and oropharynx in the study. Our follow-up period is limited with a range of 2–26 months with an actuarial median follow-up of 11 months.

**Conclusion**

Our exploratory results suggest that a reduction in MTV or TLG of 42% and 55%, respectively, before and after induction chemotherapy, predicts short-term event-free survival in patients with advanced head and neck squamous cell cancer (stages III and IV). These exploratory metabolic volumetric imaging parameters and cut points for predicting short to intermediate term outcome need validation in a larger and prospective study, which would include other important clinical parameter such as HPV status.

**Acknowledgements**

Rathan Subramaniam was supported by a GE-AUR Research fellowship and received Siemens molecular imaging and MJ Fox foundation research grants and research support from MIM Software Inc.. He is a speaker and consultant for Eli Lilly (Alzheimer’s disease).

**References**


