



# Utility and limitations of metabolic parameters in head and neck cancer: finding a practical segmentation method

Jefferson Rijo-Cedeño<sup>1,3</sup> · Jorge Mucientes<sup>2</sup> · Ithzel María Villarreal<sup>4</sup> · Ana Royuela<sup>5</sup> · Patricia García Vicente<sup>1</sup> · José Ramón García-Berrocal<sup>1,3</sup>

Received: 21 November 2021 / Accepted: 11 April 2022 / Published online: 2 May 2022  
© The Author(s) 2022

## Abstract

**Purpose** Although metabolic tumor volume (MTV) and total lesion glycolysis (TLG) have shown good prognostic value in head and neck cancer (HNC), there are still many issues to resolve before their potential application in standard clinical practice. The purpose of this study was to compare the discrimination ability of two relevant segmentation methods in HNC and to evaluate the potential benefit of adding lymph nodes' metabolism (LNM) to the measurements.

**Methods** We retrospectively analyzed a recently published database of 62 patients with HNC treated with chemoradiotherapy. MTV and TLG were measured using an absolute threshold of SUV2.5. Comparison analysis with previously published background-level threshold (BLT) results was done through Concordance index (C-index) in eight prognostic models.

**Results** BLT obtained better C-index values in five out of the eight models. The addition of LNM improved C-index values in six of the prognostic models.

**Conclusion** We found a potential benefit in adding LNM to the main tumor measurements, as well as in using a BLT for MTV segmentation compared to the most commonly used SUV2.5 threshold. Despite its limitations, this study suggests a practical and simple manner to use these parameters in standard clinical practice, aiming to help elaborate a general consensus.

**Keywords** PET · Head and neck · Cancer · MTV · TLG

## Introduction

The number of publications analyzing the prognostic value of fluorine-18-fluorodeoxyglucose (<sup>18</sup>F-FDG) metabolic tumor volume (MTV) and total lesion glycolysis (TLG) in head and neck squamous cell carcinoma (HNSCC) has been increasing in the last decades. Most of the studies find a good prognostic capability in these variables, as they provide

valuable pretreatment information of a higher risk of recurrence [1–5]. Although widely used, maximum standardized uptake value (SUV<sub>max</sub>) results are more heterogeneous, and conflicting results are found in the literature regarding its prognostic potential [1, 2, 6, 7].

With the current available data, MTV and TLG seem to have a good prognostic value in HNSCC. Nevertheless, there are still several limitations to address before they can be

✉ Jefferson Rijo-Cedeño  
jefferson.rijo@estudiante.uam.es

Jorge Mucientes  
jorge.mucientes@salud.madrid.org

Ithzel María Villarreal  
invillarrealp@gmail.com

Ana Royuela  
aroyuela@idiphim.org

Patricia García Vicente  
patriciaurria@gmail.com

José Ramón García-Berrocal  
jrgarciab@yahoo.com

<sup>1</sup> Department of Otorhinolaryngology-Head and Neck Surgery, Puerta de Hierro University Hospital, Majadahonda, Spain

<sup>2</sup> Department of Nuclear Medicine, Puerta de Hierro University Hospital, Majadahonda, Spain

<sup>3</sup> Departamento de Cirugía, Universidad Autónoma de Madrid, Madrid, Spain

<sup>4</sup> Department of Otorhinolaryngology-Head and Neck Surgery, Fuenlabrada University Hospital, Fuenlabrada, Spain

<sup>5</sup> Biostatistics Unit, Biomedical Research Institute IDIPHISA, CIBERESP, Majadahonda, Spain

implemented in standard clinical practice. Although most studies show good prognostic results, their methodologies vary considerably as there is no consensus yet on standardization of the calculations. First, studies differ on the volume of interest (VOI); some use the primary tumor as the VOI, and some add the metabolism of the affected lymph nodes to the primary tumor measurements. Second, studies differ on the segmentation methods, which directly affect the metabolic parameters calculations and reproducibility.

A wide variety of methods have been developed for the segmentation of VOIs in positron emission tomography (PET) scans. In general, these methods can be classified in two big groups: threshold-based and algorithm-based methods. Because of their practicality and simplicity, threshold-based methods are the most widely used in HNSCC studies. They define the burden of the VOIs based on a pre-established threshold using one of four different approaches: an absolute threshold (using all volumes above a certain SUV level [usually 2.5]), a relative threshold (using, in general, 40–50% of the  $SUV_{max}$  of the lesion), a background-level threshold (BLT) (using the SUV of a reference region, e.g., liver or mediastinum SUV), and adaptative thresholds (adjusting calculations to different measurable variables of the images, with several approaches to obtain a threshold). Algorithm-based methods constitute a heterogeneous and more complex group of methods that require specific software for segmentation of the VOIs [8, 9].

Aiming to compare two relevant segmentation methods, the purpose of this study was to evaluate the discrimination power of MTV and TLG delineated using an absolute threshold of SUV 2.5, as this method is the most used in the majority of HNSCC papers, in a combined analysis with the data of our recently reported results (where we segmented using a BLT with the liver as a reference region) [10]. Furthermore, we analyzed the potential benefit of adding the affected lymph nodes metabolism (LNM) to MTV and TLG measurements.

## Materials and methods

### Study population

We have analyzed the population published in a recent report [10]. After the approval of The Institutional Ethics Committee of our hospital, we conducted a retrospective examination of the oncological records of patients treated at Puerta de Hierro University Hospital, Majadahonda, Spain, between January 2012 and December 2018. We included patients over 18 years of age diagnosed with primary stage III–IV HNSCC without distant metastasis that had been treated with radiotherapy alone or concurrent chemoradiotherapy who underwent pretreatment  $^{18}F$ -FDG PET-CT

for initial staging and therapy planning with no more than 6 weeks between image acquisition and treatment initiation. Exclusion criteria implicated patients with synchronous or metachronous lesions, patients who had undergone previous treatments, and patients with nasopharyngeal cancer. Informed consent was waived because of the retrospective design.

### Radiation therapy and follow-up

Patients were immobilized with a thermoplastic head-shoulders mask and simulated in supine position. Planning 3 mm slice thickness contrast-enhanced computer tomography (CT) was obtained for all patients. These images were combined with pretreatment  $^{18}F$ -FDG PET for better delineation of treatment volumes. Clinical target volumes (CTV) were as follows: CTV1: primary and affected lymph nodes gross tumor volumes; CTV2: CTV1 and the “high risk” first uninvolved lymph node region; CTV3: elective bilateral lymph nodes concurring with international guidelines [11]. By adding 0.5 mm to the CTVs, we generated the corresponding planning target volumes (PTVs). Treatment with intensity modulated radiation therapy was administered in a TomoTherapy HiArt unit equipped with an image-guided radiation therapy system and was dispensed in 32 fractions using the simultaneous integrated boost-up to doses for PTV1, PTV2, and PTV3 of 69.12, 57.6 and 53.12 Gray (Gy), respectively. Dose-limiting constraints for organs at risk were the following: Dmax of 45 Gy for spinal cord;  $V_{28} < 50\%$  for parotid glands;  $V_{65} < 10\%$  for mandible; Dmax of 55 Gy for brain stem; and Dmean  $< 50$  Gy for constrictor muscles.

Two protocols consisting in a regimen of cisplatin (two courses of 20 mg/m<sup>2</sup>/d in days 1–4 and 29–32) plus oral tegafur (two courses of 1200 mg/d in days 1–14 and 29–43) [12], and a regimen of six cycles of weekly cisplatin (40 mg/m<sup>2</sup>) were administered as concomitant chemotherapy for most patients. Patients who were not fit for a cisplatin-based protocol were treated with Cetuximab (400 mg/m<sup>2</sup> initial dose, followed by seven weekly doses of 250 mg/m<sup>2</sup>). Patients whose physical conditions were not suitable for chemotherapy were treated with radiotherapy alone.

Two to four months after treatment completion, an  $^{18}F$ -FDG PET-CT was performed in all patients to assess adequate treatment response. Follow-up was done every 6 months for the first 3 years and once a year for 2 more years with head and neck imaging (CT or magnetic resonance imaging). Salvage surgery or palliative treatment was indicated in case of recurrent or persisting disease.

### $^{18}F$ -FDG PET-CT imaging

Prior to the  $^{18}F$ -FDG i.v. injection (approximately 350–400 MBq), patients were required to have a serum

glucose concentration below 180 mg/dL after fasting for at least 6 h. A helical multidetector CT was used to perform non-contrast-enhanced scans from the top of the skull to the mid-tight with the following parameters: 110 kVp, a maximum modulated milliamperage of 85 mAs, and six slices with a 5.0 mm thickness. CT images were then used for image fusion and attenuation correction. A multimodality Siemens Biograph 6 scan (Biograph; Siemens, Erlangen, Germany) was used to acquire the PET images after 50–60 min of  $^{18}\text{F}$ -FDG i.v. injection with the following parameters: 4-min' scan per bed position  $\times$  7–8 positions and ordered-subset expectation maximization reconstruction (four iterations, eight subsets).

### Image interpretation

$^{18}\text{F}$ -FDG PET images were analyzed by an experienced nuclear medicine practitioner (J.M.) using a Siemens Leonardo reading station. Patient's clinical outcomes were unknown during the analysis using the program Syngo®. via (Siemens Healthineers, Muenchen, Germany). To obtain semiquantitative metabolic parameters, image pixels were converted into standardized uptake values. Then, we set a SUV of 2.5 for VOI segmentation of the tumor and the affected lymph nodes. All  $^{18}\text{F}$ -FDG-avid lesions were then manually encircled in three imaging planes, and the volumes with  $\geq$  SUV2.5 were defined as the VOIs. Tumor and lymph nodes' MTV were automatically calculated by the software, and those two volumes were then analyzed individually and combined. TLG was obtained using the formula: MTV  $\times$  SUVmean.

### Statistical analysis

First, we made a descriptive analysis representing categorical variables with absolute and relative frequencies. Numerical variables were evaluated through medians with percentiles 25 and 75 (P25-P75), means with standard deviation, and minimum and maximum values.

For the survival analysis, overall survival (OS) was defined as the time between treatment initiation and death from any cause. Disease-free survival (DFS) was defined as the time between treatment initiation and relapse or death from any cause, whichever happened first. Patients who were followed-up until the end of the study with no events, as well as patients lost in the follow-up with no events were censored at the date of last contact. Kaplan–Meier curves were estimated and differences between survival curves were compared through the log-rank test. Survival analysis was made with the combination of lymph nodes and main tumor MTV and TLG.

To obtain the optimal cutoff points, receiver-operating characteristics (ROC) curves with the survival data were

developed [13], and the points were also evaluated with the Youden [14] and Liu [15] methods. To assess the association between MTV and TLG measured by SUV2.5 with OS and DFS, Cox proportional hazard models were executed. We tested each variable in two ways: as a continuous variable (per 10-ml increment for MTV and per 25-g increment for TLG) and as a dichotomized variable with the obtained cutoff points. We then developed an univariable approach, and a multivariable model adjusting by age, T-classification, and N-classification. Proportional hazards assumption was checked using Schoenfeld residuals. Hazard ratios (HR) with their corresponding 95% confidence intervals are shown.

To test the discrimination ability of the models, Harrell's/Concordance Index (C-index) was estimated. Values of C-index near 0.5 indicate that the discrimination ability is no better than chance. Values near 1 indicate that the model is good at determining which of two patients will have an event first [16].

To evaluate the effect of lymph nodes' metabolism in the discrimination ability of the variables, we created several univariable regression models for tumor, lymph nodes', and tumor + lymph nodes' MTV and TLG with both absolute threshold method (SUV2.5) and our previous published data with a background-level threshold method. [10]. To compare the latter two segmentation methods, we analyzed eight prognostic models using the combined data from both studies: MTV and TLG continuous and dichotomized models with the two segmentation methods mentioned before.

Significance level was set at 0.05. Software used was Stata 16 (StataCorp. 2019. *Stata Statistical Software: Release 16*. College Station, TX: StataCorp LLC.)

## Results

### Demographics

Seventy-nine oncological records were assessed from patients diagnosed with stage III and IV HNSCC. Two patients had a longer than 6 weeks interval between treatment initiation and PET acquisition and thus were excluded. Pre-treatment  $^{18}\text{F}$ -FDG PET-TC was performed in a different center in 15 patients and were also excluded. Therefore, 62 records were fully assessed for the analysis. The median age was 65 years (range, 38–87 years) and the female/male distribution was 13/49. Forty-nine patients were smokers. Disease locations were: oral cavity (OC): 7 patients, hypopharynx (HP): 6 patients, oropharynx (OP): 16 patients (7 human papilloma virus [HPV] positive, 6 HPV negative, and not stated HPV status in 3 patients), and larynx (L): 33 patients.

T-classification findings were as follows: T1: 2 patients; T2: 16 patients; T3: 29 patients; and T4: 15 patients. N-classification included: N0: 14 patients; N1: 5 patients; N2: 37

patients; and N3: 6 patients. Twenty-two patients had a AJCC stage III disease, 29 had a stage IVa, and 11 had a stage IVb. Most patients received treatment with concomitant chemoradiotherapy. Thirty-one of them were treated with a chemotherapy protocol of cisplatin + tegafur (OC: 4, OP: 6, HP: 4, and L: 17), 17 with a weekly cisplatin protocol (OC: 1, OP: 5, HP: 1, and L: 10), 11 with a cetuximab protocol (OC: 2, OP: 2, HP: 1, and L: 6), and 3 patients were treated with radiotherapy alone (OP: 3). Two patients died during treatment and, because of a severe case of radiodermatitis, one patient refused to complete the treatment.

### PET parameters' determination

Respectively, the median values for TLG and MTV were 176.92 g (range 5.5–2,567.04 g) and 29.94 ml (range 1.66–220.87 ml). Cutoff values estimated at a 3-years' time point for TLG and MTV were acquired using a time-dependent AUC-ROC. The same values were obtained using the Youden and Liu methods. The cutoff points were 284 g for TLG (AUC=0.68) and 60 ml for MTV (AUC=0.70). Subsequently, Cox regression analyses were performed dichotomizing the corresponding cutoff values.

### Survival analysis

The median follow-up was 59.3 months (95% CI; 48.6–68.5 months). Locoregional recurrence occurred in 21 patients and distant metastasis was found in 13 patients. Four-year DFS was 57% (95% CI 43–69). Four-year overall survival (OS) was 63% (95% CI 49–74). Global 4-year OS and DFS per disease location was, respectively, as follows: OC: 50% (95% CI=11–80) and 65% (95% CI=35–84); OP: 65% (95% CI=35–84) and 48% (95% CI=11–80); HP: 50% (95% CI=11–80) and 20% (95% CI=05–75); L: 68% (95% CI=48–81) and 60% (95% CI=41–75).

OS and DFS of larynx patients per stage were, respectively, as follows: Stage III, 75 (95% CI=40–91) and 73 (95% CI=37–90); Stage IVa, 64 (95% CI=35–82) and 54 (95% CI=28–74); Stage IVb, 50 (95% CI=06–91) and 50 (95% CI=0.6–91). OS and DFS of oropharynx patients per stage were, respectively, as follows: Stage IVa, 48 (95% CI=0.7–81) and 38 (95% CI=06–71); Stage IVb, 75 (95% CI=12–96) and 75 (95% CI=12–96). It was only possible to estimate the overall survival for patients with stage III OP cancer with an OS of 80 (95% CI=20–97). OS and DFS of oral cavity patients in stage IVb were 25 (95% CI=0.8–67) and 25 (95% CI=0.8–67). OS and DFS of hypopharynx patients in stage III were 33 (95% CI=0.9–77) and 33 (95% CI=0.9–77). It was not possible to estimate survival in stages III nor IVa OC patients nor in stages IVa and IVb of HP patients because of the absence of events during the

follow-up. DFS of stage III OP cancer was not estimable as well.

Primary local control was achieved in 12 (75%), 4 (57%) 3 (50%) and 24 (73%) of the patients with OP, OC, HP, and L disease, respectively. Primary local–regional control was achieved in 11 (69%), 4 (57%), 3 (50%), and 23 (70%) of the patients with OP, OC, HP, and L disease, respectively. A broader detailed description of the study population is described in our previous report [10].

There was a significant association with OS in both dichotomized [hazard ratio (HR)=2.61, 95% CI=1.18–5.81,  $p=0.018$ ] and continuous (HR=1.11, 95% CI=1.04–1.19,  $p=0.002$ ) MTV models in the univariable analysis. Dichotomized (HR=2.48, 95% CI=1.13–5.45,  $p=0.024$ ) and continuous (HR=1.03, 95% CI=1.01–1.05,  $p=0.001$ ) TLG models also had a significant association with OS. There was also a significant association with DFS in the dichotomized (HR=3.06, 95% CI=1.38–6.76,  $p=0.006$ ) and continuous (HR=1.16, 95% CI=1.08–1.25,  $p<0.001$ ) MTV models, as well as in the dichotomized (HR=2.87, 95% CI=1.23–6.35,  $p=0.009$ ) and continuous (HR=1.03, 95% CI=1.01–1.05,  $p<0.001$ ) TLG models in the univariable analysis (Table 1).

Both MTV and TLG models remained as independent prognostic factors for DFS and OS in the multivariable analyses. Patients with a >284 g TLG had a higher hazard for recurrence (HR=2.91, 95% CI=1.28–6.64,  $p=0.011$ ) and death (HR=2.86, 95% CI=1.25–6.48,  $p=0.013$ ). The HR for every 25-g increment of TLG was 1.03 (95% CI=1.01–1.05,  $p=0.001$ ) for DFS, and 1.03 (95% CI=1.01–1.05,  $p=0.001$ ) for OS. Patients with >60 ml MTV also had a higher hazard for recurrence (HR=3.07, 95% CI=1.36–6.94,  $p=0.007$ ) and death (HR=3.06, 95% CI=1.31–7.12,  $p=0.010$ ). HR for every 10-ml increment of MTV was 1.16 (95% CI=1.71–1.26,  $p=<0.001$ ) for DFS, and 1.12 (95% CI=1.04–1.22,  $p=0.002$ ) for OS (Table 1).

Kaplan–Meier's curves for DFS and OS using the dichotomized models of MTV and TLG are shown in Fig. 1.

### Comparison analysis

Univariable analyses per unit increment of lymph nodes', tumor, and the combination of tumor and lymph nodes' MTV found a significant association with OS and DFS in both background-level and absolute SUV2.5 segmentation methods (Table 2). The same happened with TLG, except for tumor TLG measured with the BLT method, which did not have a significant association with OS nor DFS (Table 3). The determination of tumor + lymph nodes' TLG obtained better C-index values than the determination of tumor TLG alone with both segmentation methods. The same happened with MTV measurements using the absolute SUV2.5 method. On the contrary, C-index values were lower with

**Table 1** Cox regression analyses for MTV and TLG segmented with SUV2.5

	Overall survival				Disease-free survival			
	Univariable analysis		Multivariable analysis <sup>a</sup>		Univariable analysis		Multivariable analysis <sup>†</sup>	
	HR (95% CI)	<i>p</i> -value	HR (95% CI)	<i>p</i> -value	HR (95% CI)	<i>p</i> -value	HR (95% CI)	<i>p</i> -value
Dichotomized MTV model		0.018		0.010		0.006		0.007
≤ 60 ml	1		1		1		1	
> 60 ml	2.61 (1.18–5.81)		3.06 (1.31–7.12)		3.06 (1.38–6.76)		3.07 (1.36–6.94)	
Continuous MTV model		0.002		0.002		< 0.001		< 0.001
Per 10-ml increment	1.11 (1.04–1.19)		1.12 (1.04–1.22)		1.16 (1.08–1.25)		1.16 (1.71–1.26)	
Dichotomized TLG model		0.024		0.013		0.009		0.011
≤ 284 g	1		1		1		1	
> 284 g	2.48 (1.13–5.45)		2.86 (1.25–6.48)		2.87 (1.23–6.35)		2.91 (1.28–6.64)	
Continuous TLG model		0.001		0.001		< 0.001		0.001
Per 25-g increment	1.03 (1.01–1.05)		1.03 (1.01–1.05)		1.03 (1.01–1.05)		1.03 (1.01–1.05)	

<sup>a</sup>Adjusted by age, T-classification, and N-classification

HR hazard ratio, CI confidence interval, MTV metabolic tumor volume, TLG total lesion glycolysis, C-index Harrell's C-index, ml milliliter, g gram

the addition of LNM to tumor MTV measurements with the background-level method for both OS and DFS (Tables 2–3).

Among the eight models, we created comparing the two segmentation methods, BLT had a better C-index in five of them. The four dichotomized models performed better with the BLT method. When comparing the continuous models, SUV 2.5 models worked better to predict the DFS. When predicting OS, only the continuous model of SUV 2.5 MTV had better C-index values than the BLT (Table 4).

## Discussion

Although MTV and TLG have consistently shown good prognostic value in patients with HNSCC in many studies [1–5], there are still several factors limiting their application in standard clinical practice. Because of the variability in the methodology used in the different publications, it is not possible to establish adequate nor generalizable cutoff values. For their correct implementation, it is necessary to firstly standardize one segmentation method among the many available ones, as well as establishing whether we should incorporate the metabolism of the affected lymph nodes in the measurements or not.

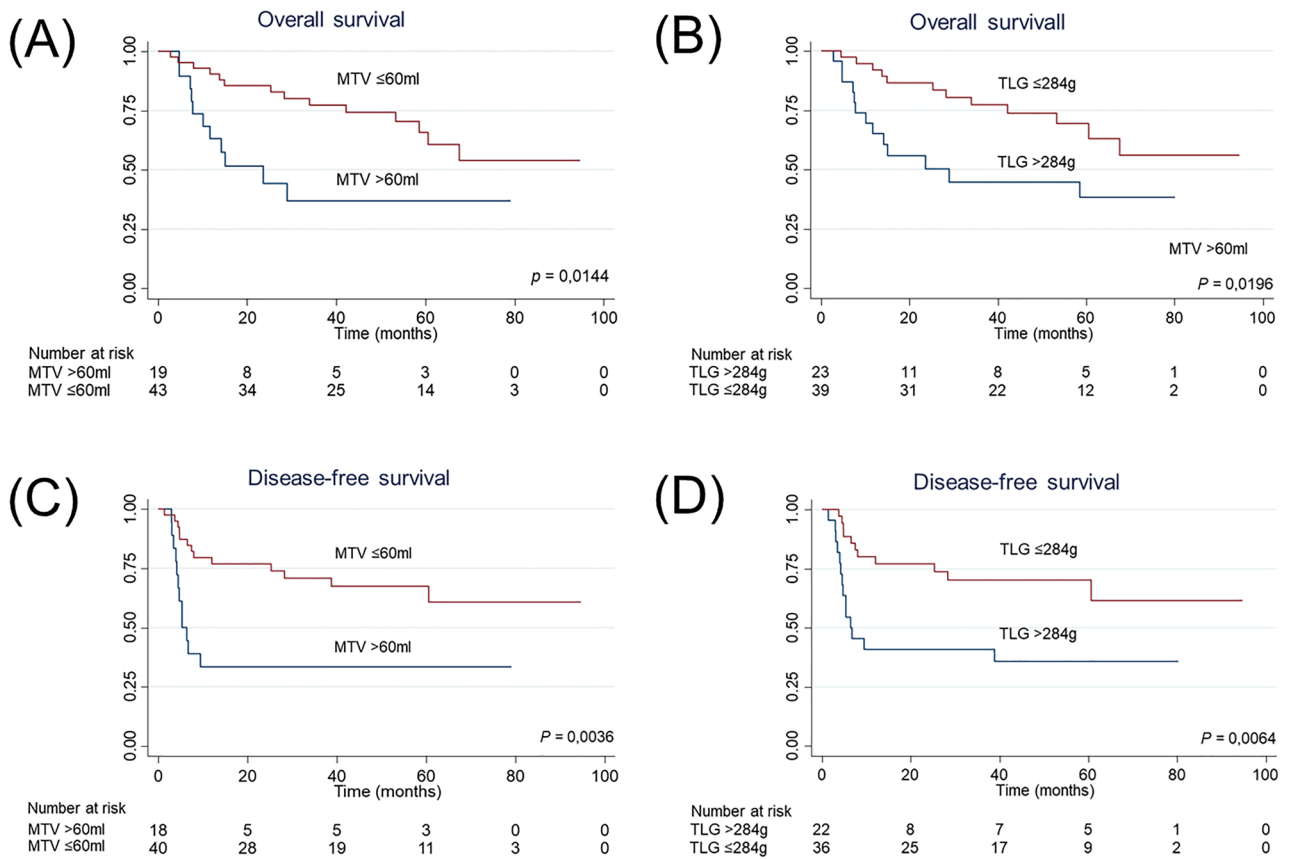
In this study, we evaluated the prognostic value of MTV and TLG delineated using an absolute threshold of SUV 2.5 in our recently reported database (where we segmented

using a BLT with the liver as a reference region) [10]. Subsequently, we created eight prognostic models using the results of both studies to assess the discrimination power of the two methods through their C-index values. Our results found an independent statistical association with survival for both methods in this population (Table 1), and the BLT obtained better C-Index values in five out of the eight prognostic models, including all dichotomized models (Table 4).

We have found several studies comparing threshold and algorithms' segmentation methods, and the latter ones seem to perform better at discriminating patients at risk of recurrence [17–22]. It is to be expected that more complex and specific methods would provide a better prognosis prediction. However, considering the many different algorithms available and the limited studies found from each of them, it is difficult to imagine any of these methods having a widespread application at the moment. The same happens with adaptive thresholds, where there is no consensus yet on which factors to consider for the calculations [8]. Thus, threshold methods seem to be the most practical and reproducible segmentation methods with current PET scanners, and they have systematically provided good prognostic information in HNSCC.

Among the five studies comparing fixed absolute, fixed relative, and background-level threshold methods we have found, four of them obtained a better result with the latter, concurring with our results [19, 23–25]. One study did





**Fig. 1** Kaplan–Meier’s curves for DFS and OS using the MTV and TLG dichotomized models. A, Curves for OS and a cutoff of 60 ml of MTV. B, Curves for OS and a cutoff of 284 g of TLG; C, curves for

DFS and a cutoff of 60 ml of MTV. D, Curves for DFS and a cutoff of 284 g of TLG. *MTV* metabolic tumor volume, *TLG* total lesion glycolysis, *ml* milliliter, *g* gram

**Table 2** Univariable Cox regression analyses for the volumes of interest of MTV

	Overall survival			Disease-free survival		
	HR (95% CI) <sup>a</sup>	<i>p</i> -value	C-index	HR (95% CI) <sup>a</sup>	<i>p</i> -value	C-index
T MTV with SUV2.5 threshold	1.02 (1.01–1.03)	0.009	0.656	1.02 (1.01–1.03)	0.002	0.695
N MTV with SUV2.5 threshold	1.01 (1.003–1.027)	0.013	0.586	1.01 (1.003–1.02)	0.012	0.587
T+N with MTV SUV2.5 threshold	1.01 (1.01–1.02)	0.002	0.667	1.01 (1.01–1.02)	<0.001	0.703
T MTV with background-level threshold	1.02 (1.01–1.04)	0.005	0.659	1.03 (1.01–1.0)	0.001	0.688
N MTV with background-level threshold	1.02 (1.004–1.04)	0.011	0.575	1.01 (1.002–1.03)	0.025	0.574
T+N MTV with background-level threshold	1.02 (1.01–1.03)	0.002	0.656	1.02 (1.01–1.03)	<0.001	0.681

<sup>a</sup>Per unit increment

*HR* hazard ratio, *CI* confidence interval, *MTV* metabolic tumor volume, *C-index* Harrell’s C-index, *SUV* standardized uptake value, *T* tumor, *N* lymph node

not find a statistical significance result for the background-level method [26]. Absolute and relative threshold methods are dependent of the standardized uptake value (SUV) and, therefore, are directly affected by SUV variations. Patients’ preparation and characteristics, PET scanner quality and differences in image analysis protocols have a direct impact in SUV calculations [27–29]. To reduce this variability, some

authors suggest to either use a background-level threshold [24, 28, 30, 31] or a gradient-based method [18, 20, 21].

To analyze the potential benefit of adding lymph node’s metabolism to the measurements, we created eight prognostic models in a combined analysis (Tables 2–3). We found an association between tumor and lymph nodes’ MTV and TLG with survival in six of them. Only tumor TLG using the

**Table 3** Univariable Cox regression analyses for the volumes of interest of TLG

	Overall survival			Disease-free survival		
	HR (95% CI) <sup>a</sup>	<i>p</i> -value	C-index	HR (95% CI) <sup>a</sup>	<i>p</i> -value	C-index
T TLG with SUV2.5 threshold	1.001 (1.004–1.003)	0.009	0.652	1.003 (1.001–1.004)	0.001	0.693
N TLG with SUV2.5 threshold	1.002 (1.001–1.003)	0.005	0.586	1.001 (1.001–1.002)	0.014	0.586
T+N with TLG SUV2.5 threshold	1.001 (1.001–1.002)	0.001	0.660	1.001 (1.001–1.002)	<0.001	0.706
T TLG with background-level threshold	1.001 (1.01–1.001)	0.281	0.629	1.001 (1.01–1.001)	0.245	0.671
N TLG with background-level threshold	1.002 (1.001–1.003)	0.006	0.545	1.001 (1.001–1.002)	0.020	0.566
T+N TLG with background-level threshold	1.001 (1.001–1.002)	0.001	0.649	1.001 (1.001–1.002)	<0.001	0.694

<sup>a</sup>Per unit increment

HR hazard ratio, CI confidence interval, TLG total lesion glycolysis, C-index Harrell's C-index, SUV standardized uptake value, T tumor, N lymph node

**Table 4** C-index comparison analysis with multivariable models of the analyzed segmentation methods

Model	C-index values for overall survival		C-index values for disease-free survival	
	Absolute threshold of SUV2.5	Background-level threshold <sup>a</sup>	Absolute threshold of SUV2.5	Background-level threshold <sup>a</sup>
Dichotomized MTV model	0.664	0.673	0.657	0.673
Continuous MTV model	0.681	0.676	0.677	0.676
Dichotomized TLG model	0.676	0.696	0.669	0.675
Continuous TLG model	0.662	0.663	0.678	0.677

<sup>a</sup>Using the liver as the reference region

MTV metabolic tumor volume, TLG total lesion glycolysis, C-index Concordance index, ml milliliter, g gram

BLT method did not have a statistical significance for neither DFS nor OS. After the addition of the LNM to the main tumor measurements, we obtained better C-index values in six of the models, and BLT tumor TLG gained statistical significance for DFS and OS. Only tumor MTV using the BLT method got lower values of C-index after the addition of the LNM. These results exhibit a possible prognostic benefit in the combination of both measurements.

In the literature, besides Castelli et al. [32] and this study, most papers do not find a statistical association between LNM and survival in HNSCC [33–40], although it may play a role in patients with HPV + oropharyngeal cancer [41–44] and in hypopharyngeal patients with a low MTV [35]. Hoshikawa et al. [45] share the same findings we obtained in our BLT tumor TLG models, where statistical significance was achieved only after the addition of the LNM. To the best of our knowledge, this is the first study testing the possible benefit of combining both tumor and lymph node's measurements in HNSCC.

This study has numerous limitations. The potential biases of a retrospective data compilation in a reduced

single-center population with different disease locations limit the generalization of the results. Although there was a tendency of a better performance of the BLT method, the difference in the C-index values were millimetric. Even so, and with all due limitations, this study addresses major concerns in the implementation of MTV and TLG in today's clinical practice, hoping to help elaborate a consensus on how to measure these variables.

Although PET's metabolic parameters seem to offer valuable prognostic information and, perhaps in the future, could aid in treatment selection [46] and in dynamic radiation dose modification [47], more prospective and multicentric studies with rigorous standardization in PET preparation are still needed to obtain quality evidence regarding how and when to use these parameters in HNSCC. Our results found a possible statistical benefit in adding lymph nodes metabolism to the measurements, as well as in using a BLT for segmentation compared to the most used SUV2.5 threshold. Considering this method only adds no more than a minute of work and can be measured with current PET scanners with no specific software, it may be

a suitable choice to evaluate MTV and TLG in HNSCC patients.

## Conclusions

Out of the eight prognostic models created, the BLT method had a better prediction power than SUV2.5 in five of them, and the addition of the affected lymph node's metabolism to MTV and TLG measurements improved their prediction power in six of the eight models. Despite the limitations of this study, our results suggest a practical and simple manner on which these metabolic parameters can be implemented in today's clinical practice.

**Acknowledgements** This study did not require any financial support.

**Author contribution** JR-C: literature review; analysis and interpretation of data; drafting of the manuscript; final approval of the manuscript. JM: data collection; analysis and interpretation of data; final approval of the manuscript. IMV: data collection; analysis and interpretation of data; final approval of the manuscript. AR: statistical analysis; final approval of the manuscript. PGV: data collection; analysis and interpretation of data; final approval of the manuscript. JRG-B: research coordinator; final approval of the manuscript.

**Funding** Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

## Declarations

**Conflict of interest** The authors declare no conflicts of interest that could be perceived as prejudicing the impartiality of the research reported. This study did not receive any financial support and it has been approved by the Institutional Ethics Committee of Puerta de Hierro University Hospital. Because of the retrospective design of this study, informed consent was waived.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Bonomo P, Merlotti A, Olmetto E et al (2018) What is the prognostic impact of FDG PET in locally advanced head and neck squamous cell carcinoma treated with concomitant chemo-radiotherapy? A systematic review and meta-analysis. *Eur J Nucl Med Mol Imaging* 45(12):2122–2138. <https://doi.org/10.1007/s00259-018-4065-5>
- Pak K, Cheon GJ, Nam H-Y et al (2014) Prognostic value of metabolic tumor volume and total lesion glycolysis in head and neck cancer: a systematic review and meta-analysis. *J Nucl Med* 55(6):884–890. <https://doi.org/10.2967/jnumed.113.133801>
- Xie K, Chen J, Zou J, Chen L, Gong L. Tumor volumes predict prognosis in head and neck cancer: a meta-analysis. *Trans Cancer Res.* 2017;6(4):687–697. <https://doi.org/10.21037/15151>
- Rijo-Cedeño J, Mucientes J, Álvarez O et al (2020) Metabolic tumor volume and total lesion glycolysis as prognostic factors in head and neck cancer: Systematic review and meta-analysis. *Head Neck* 42(12):3744–3754. <https://doi.org/10.1002/hed.26461>
- Creff G, Devillers A, Depeursinge A et al (2020) Evaluation of the prognostic value of FDG PET/CT Parameters for patients with surgically treated head and neck cancer: a systematic review. *JAMA Otolaryngol Head Neck Surg* 146(5):471–479. <https://doi.org/10.1001/JAMAOTO.2020.0014>
- Pak K, Cheon GJ, Kang KW, Chung JK, Kim EE, Lee DS (2015) Prognostic value of SUVmean in oropharyngeal and hypopharyngeal cancers: comparison with SUVmax and other volumetric parameters of 18F-FDG PET. *Clin Nucl Med* 40(1):9–13. <https://doi.org/10.1097/RLU.0000000000000613>
- Cacicedo J, Fernandez I, del Hoyo O et al (2017) Prognostic value of maximum standardized uptake value measured by pre-treatment 18F-FDG PET/CT in locally advanced head and neck squamous cell carcinoma. *Clin Transl Oncol* 19(11):1337–1349. <https://doi.org/10.1007/s12094-017-1674-6>
- Im H-J, Bradshaw T, Solaiyappan M, Cho SY (2018) Current methods to define metabolic tumor volume in positron emission tomography: Which one is better? *Nucl Med Mol Imaging* 52(1):5–15. <https://doi.org/10.1007/s13139-017-0493-6>
- Schinagl DAX, Vogel Wv., Hoffmann AL, van Dalen JA, Oyen WJ, Kaanders JHAM. Comparison of five segmentation tools for 18F-fluoro-deoxy-glucose-positron emission tomography-based target volume definition in head and neck cancer. *Int J Radiat Oncol Biol Phys.* 2007;69(4):1282–1289. <https://doi.org/10.1016/j.ijrobp.2007.07.2333>
- Rijo-Cedeño J, Mucientes J, Marcos SS et al (2021) Adding value to tumor staging in head and neck cancer: The role of metabolic parameters as prognostic factors. *Head Neck.* 43(8):2477–2487. <https://doi.org/10.1002/HED.26725>
- Grégoire V, Ang K, Budach W, et al. (2014) Delineation of the neck node levels for head and neck tumors: A 2013 update. *DAHANCA, EORTC, HKNPCSG, NCIC CTG, NCRI, RTOG, TROG consensus guidelines.* *Radiother Oncol.* 110(1):172–181. <https://doi.org/10.1016/j.radonc.2013.10.010>
- de La Torre A, Romero J, Montero A et al (2008) Radiochemotherapy with cisplatin and oral tegafur in advanced head and neck cancer: long-term results of a phase II study. *Tumori J* 94(4):453–458. <https://doi.org/10.1177/030089160809400402>
- Heagerty PJ, Lumley T, Pepe MS (2000) Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* 56(2):337–344. <https://doi.org/10.1111/j.0006-341X.2000.00337.x>
- Fluss R, Faraggi D, Reiser B (2005) Estimation of the Youden Index and its associated cutoff point. *Biom J* 47(4):458–472. <https://doi.org/10.1002/bimj.200410135>
- Library WO, Liu X, Jin Z (2014) Optimal survival time-related cut-point with censored data. *Stat Med* 34(3):515–524. <https://doi.org/10.1002/sim.6360>
- Kozioł JA, Jia Z (2009) The Concordance index C and the mann-whitney parameter  $Pr(X>Y)$  with randomly censored data. *Biom J* 51(3):467–474. <https://doi.org/10.1002/BIMJ.200800228>



17. Tyłski P, Stute S, Grotus N et al (2010) Comparative assessment of methods for estimating tumor volume and standardized uptake value in 18F-FDG PET. *J Nucl Med* 51(2):268–276. <https://doi.org/10.2967/jnumed.109.066241>
18. Dibble EH, Alvarez ACL, Truong MT, Mercier G, Cook EF, Subramaniam RM (2012) 18F-FDG metabolic tumor volume and total glycolytic activity of oral cavity and oropharyngeal squamous cell cancer: Adding value to clinical staging. *J Nucl Med* 53(5):709–715. <https://doi.org/10.2967/jnumed.111.099531>
19. Burger IA, Casanova R, Steiger S et al (2016) 18F-FDG PET/CT of non-small cell lung carcinoma under neoadjuvant chemotherapy: Background-based adaptive-volume metrics outperform TLG and MTV in predicting histopathologic response. *J Nucl Med* 57(6):849–854. <https://doi.org/10.2967/jnumed.115.167684>
20. Geets X, Lee JA, Bol A, Lonneux M, Grégoire V (2007) A gradient-based method for segmenting FDG-PET images: methodology and validation. *Euro J Nucl Med Mol Imaging* 34(9):1427–1438. <https://doi.org/10.1007/s00259-006-0363-4>
21. Wanet M, Lee JA, Weynand B et al (2011) Gradient-based delineation of the primary GTV on FDG-PET in non-small cell lung cancer: A comparison with threshold-based approaches, CT and surgical specimens. *Radiother Oncol* 98(1):117–125. <https://doi.org/10.1016/j.radonc.2010.10.006>
22. Zaidi H, Abdoli M, Fuentes CL, el Naqa IM (2012) Comparative methods for PET image segmentation in pharyngolaryngeal squamous cell carcinoma. *Eur J Nucl Med Mol Imaging* 39(5):881–891. <https://doi.org/10.1007/s00259-011-2053-0>
23. Moon SH, Choi JY, Lee HJ et al (2013) Prognostic value of 18F-FDG PET/CT in patients with squamous cell carcinoma of the tonsil: comparisons of volume-based metabolic parameters. *Head Neck* 35(1):15–22. <https://doi.org/10.1002/hed.22904>
24. Hofheinz F, Bütof R, Apostolova I, et al. An investigation of the relation between tumor-to-liver ratio (TLR) and tumor-to-blood standard uptake ratio (SUR) in oncological FDG PET. *EJNMMI Res.* 2016;6(1). <https://doi.org/10.1186/s13550-016-0174-y>
25. Burger IA, Vargas HA, Apte A et al (2014) PET quantification with a histogram derived total activity metric: Superior quantitative consistency compared to total lesion glycolysis with absolute or relative SUV thresholds in phantoms and lung cancer patients. *Nucl Med Biol* 41(5):410–418. <https://doi.org/10.1016/j.nucmedbio.2014.02.006>
26. Yoo IR, Chung SK, Park HL, et al. (2014) Prognostic value of SUVmax and metabolic tumor volume on 18F-FDG PET/CT in early stage non-small cell lung cancer patients without LN metastasis. *Biomed Mater Eng.* 24. IOS Press; 3091–3103. <https://doi.org/10.3233/BME-141131>
27. Mah K, Caldwell CB (2008) Biological Target Volume. In: Paulino AC, Teh BS, eds. *PET-CT in Radiotherapy Treatment Planning*. Elsevier Inc.:52–89. <https://doi.org/10.1016/B978-1-4160-3224-3.50007-4>
28. Wahl RL, Jacene H, Kasamon Y, Lodge MA (2009) From RECIST to PERCIST: Evolving considerations for PET response criteria in solid tumors. *J Nucl Med.* 50(SUPPL. 1). <https://doi.org/10.2967/jnumed.108.057307>
29. Paquet N, Albert A, Foidart J, Hustinx R (2004) Within-patient variability of (18)F-FDG: standardized uptake values in normal tissues. *J Nucl Med* 45(5):748–788
30. Cheson BD, Pfistner B, Juweid ME et al (2007) Revised response criteria for malignant lymphoma. *J Clin Oncol* 25(5):579–586. <https://doi.org/10.1200/JCO.2006.09.2403>
31. Boktor RR, Walker G, Stacey R, Gledhill S, Pitman AG (2013) Reference range for intrapatient variability in blood-pool and liver SUV for 18F-FDG PET. *J Nucl Med* 54(5):677–682. <https://doi.org/10.2967/jnumed.112.108530>
32. Castelli J, Depeursinge A, Ndoh V et al (2017) A PET-based nomogram for oropharyngeal cancers. *Eur J Cancer* 75:222–230. <https://doi.org/10.1016/j.ejca.2017.01.018>
33. Koyasu S, Nakamoto Y, Kikuchi M et al (2014) Prognostic value of pretreatment 18F-FDG PET/CT parameters including visual evaluation in patients with head and neck squamous cell carcinoma. *AJR Am J Roentgenol* 202(4):851–858. <https://doi.org/10.2214/AJR.13.11013>
34. Miyabe J, Hanamoto A, Tatsumi M et al (2017) Metabolic tumor volume of primary tumor predicts survival better than T classification in the larynx preservation approach. *Cancer Sci* 108(10):2030–2038. <https://doi.org/10.1111/cas.13345>
35. Okazaki E, Kawabe J, Oishi M et al (2019) Prognostic significance of pretreatment 18F-fluorodeoxyglucose positron emission tomography evaluation using metabolic tumor volume of the primary tumor and lymph nodes in advanced hypopharyngeal cancer. *Head Neck* 41(3):739–747. <https://doi.org/10.1002/hed.25429>
36. Ryu IS, Kim JS, Roh JL et al (2014) Prognostic significance of preoperative metabolic tumour volume and total lesion glycolysis measured by 18F-FDG PET/CT in squamous cell carcinoma of the oral cavity. *Eur J Nucl Med Mol Imaging* 41(3):452–461. <https://doi.org/10.1007/s00259-013-2571-z>
37. Kim K-R, Shim H-J, Hwang J-E et al (2018) The role of interim FDG PET-CT after induction chemotherapy as a predictor of concurrent chemoradiotherapy efficacy and prognosis for head and neck cancer. *Eur J Nucl Med Mol Imaging* 45(2):170–178. <https://doi.org/10.1007/s00259-017-3836-8>
38. Lin Y-C, Chen S-W, Hsieh T-C et al (2015) Risk stratification of metastatic neck nodes by CT and PET in patients with head and neck cancer receiving definitive radiotherapy. *J Nucl Med* 56(2):183–189. <https://doi.org/10.2967/jnumed.114.148023>
39. Zhang H, Seikaly H, Nguyen N-T et al (2016) Validation of metabolic tumor volume as a prognostic factor for oral cavity squamous cell carcinoma treated with primary surgery. *Oral Oncol* 57:6–14. <https://doi.org/10.1016/j.oraloncology.2016.03.013>
40. Lin P, Min M, Lee M et al (2017) Nodal parameters of FDG PET/CT performed during radiotherapy for locally advanced mucosal primary head and neck squamous cell carcinoma can predict treatment outcomes: SUVmean and response rate are useful imaging biomarkers. *Eur J Nucl Med Mol Imaging* 44(5):801–811. <https://doi.org/10.1007/s00259-016-3584-1>
41. Alluri KC, Tahari AK, Wahl RL, Koch W, Chung CH, Subramaniam RM (2014) Prognostic value of FDG PET metabolic tumor volume in human papillomavirus-positive stage III and IV oropharyngeal squamous cell carcinoma. *AJR Am J Roentgenol* 203(4):897–903. <https://doi.org/10.2214/AJR.14.12497>
42. Kim KH, Lee J, Chang JS et al (2016) Prognostic value of FDG-PET volumetric parameters in patients with p16-positive oropharyngeal squamous cell carcinoma who received curative resection followed by postoperative radiotherapy or chemoradiotherapy. *Head Neck* 38(10):1515–1524. <https://doi.org/10.1002/hed.24465>
43. Tahari AK, Alluri KC, Quon H, Koch W, Wahl RL, Subramaniam RM (2014) FDG PET/CT imaging of oropharyngeal squamous cell carcinoma: characteristics of human papillomavirus-positive and -negative tumors. *Clin Nucl Med* 39(3):225–231. <https://doi.org/10.1097/RLU.0000000000000255>
44. Kikuchi M, Koyasu S, Shinohara S et al (2015) Prognostic value of pretreatment 18F-fluorodeoxyglucose positron emission tomography/CT volume-based parameters in patients with oropharyngeal squamous cell carcinoma with known p16 and p53 status. *Head Neck* 37(10):1524–1531. <https://doi.org/10.1002/hed.23784>
45. Hoshikawa H, Mori T, Maeda Y et al (2017) Influence of volumetric 4'-[methyl-11C]-thiothymidine PET/CT parameters for prediction of the clinical outcome of head and neck cancer patients. *Ann Nucl Med* 31(1):63–70. <https://doi.org/10.1007/s12149-016-1131-6>

46. Yabuki K, Sano D, Shiono O et al (2017) Surgery-based versus radiation-based treatment strategy for a high metabolic volume laryngeal cancer. *Laryngoscope* 127(4):862–867. <https://doi.org/10.1002/lary.26233>
47. Piperdi H, Portal D, Neibart SS, Yue NJ et al (2021) Adaptive Radiation Therapy in the Treatment of Lung Cancer: An Overview of the Current State of the Field. *Front Oncol* 11:770382. <https://doi.org/10.3389/fonc.2021.770382>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.