# Metabolic tumor volume of primary tumor predicts survival better than T classification in the larynx preservation approach

Junji Miyabe,[1] Atsushi Hanamoto,[1] Mitsuaki Tatsumi,[2] Toshimitsu Hamasaki,[3] (iD) Yukinori Takenaka,[1] Susumu Nakahara,[1] Toshihiro Kishikawa,[1] Motoyuki Suzuki,[1] Norihiko Takemoto,[1] Takahiro Michiba,[1] Yasuo Yoshioka,[4] Fumiaki Isohashi,[4] Koji Konishi,[4] Kazuhiko Ogawa,[4] Jun Hatazawa[2] and Hidenori Inohara[1] (iD)

Departments of [1]Otorhinolaryngology–Head and Neck Surgery; [2]Nuclear Medicine and Tracer Kinetics, Osaka University Graduate School of Medicine; [3]Research and Development Initiative Center, National Cerebral and Cardiovascular Center; [4]Department of Radiation Oncology, Osaka University Graduate School of Medicine, Osaka, Japan

We aimed to determine whether pretreatment metabolic tumor volume of the primary tumor (T-MTV) or T classification would be a better predictor of laryngectomy-free survival (LFS) and overall survival (OS) after chemoradiotherapy in patients with locally advanced laryngeal or hypopharyngeal cancer requiring total laryngectomy. We analyzed 85 patients using a Cox proportional hazards model and evaluated its usefulness by Akaike's information criterion. A T-MTV cut-off value was determined by time-dependent receiver operating characteristic curve analysis. Interobserver reliability for measuring T-MTV was estimated by the intraclass correlation coefficient (ICC). After adjustment for covariables, T-MTV, irrespective of whether a continuous or dichotomized variable, and T classification remained independent predictors of LFS and OS. Large T-MTV (>28.7 mL) was associated with inferior LFS (hazard ratio [HR], 4.16; 95% confidence interval [CI], 1.97–8.70; $P$ = 0.0003) and inferior OS (HR, 3.18; 95% CI, 1.47–6.69; $P$ = 0.004) compared with small T-MTV (≤28.7 mL). The T-MTV model outperformed the T classification model in predicting LFS and OS ($P$ = 0.007 and 0.01, respectively). Three-year LFS and OS rates for patients with small versus large T-MTV were 68% vs 9% ($P$ < 0.0001) and 77% vs 25% ($P$ < 0.0001), respectively, whereas those for patients with T2-T3 versus T4a were 61% vs 31% ($P$ = 0.003) and 71% vs 48% ($P$ = 0.10), respectively. ICC was 0.99 (95% CI, 0.99–1.00). Given the excellent interobserver reliability, T-MTV is better than T classification to identify patients who would benefit from the larynx preservation approach.

Locally advanced laryngeal and hypopharyngeal cancer share the treatment goal of larynx preservation. The standard approach of larynx preservation is CRT, whereas guidelines recommend upfront total laryngectomy for T4a disease instead of CRT for fear of compromising survival.[1,2] In fact, the National Cancer Database in the USA reported that patients with T4a laryngeal cancer survived better when treated with total laryngectomy than with CRT.[3]

According to the seventh edition of the AJCC,[4] the definition of T4a laryngeal cancer includes such criteria as invasion through the outer cortex of the thyroid cartilage and/or extra-laryngeal spread, whereas the definition of T4a hypopharyngeal cancer includes inner thyroid cartilage cortex invasion. These criteria are often fundamental determinants when classifying a tumor as T4a. Although CT plays a pivotal role in evaluating thyroid cartilage invasion and extra-laryngeal spread, CT imaging has clear limitations. When CT findings were correlated with pathological findings as the gold standard, positive predictive value of CT findings for outer thyroid cartilage cortex invasion, inner thyroid cartilage cortex invasion, and extra-laryngeal

spread were 74%, 35%, and 81%, respectively.[5] Moreover, there is a considerable interobserver variation in discerning thyroid cartilage invasion and extra-laryngeal spread.[6] Accordingly, T classification based on CT findings is somewhat less objective, especially for T4a. A more objective imaging modality that would allow precise risk stratification of patients requiring total laryngectomy is mandatory.

^{18}F-FDG PET/CT, which measures and visualizes metabolic activity, represents a candidate for that imaging modality. Among quantitative PET measures, volumetric parameters such as MTV, which reflects the tumor burden with increased FDG uptake, are expected to be associated with clinical outcomes. We showed previously that T-MTV was an independent predictor of local response to CRT in laryngeal and hypopharyngeal cancer.[7]

We assumed that T-MTV would have an advantage over T classification for predicting LFS and OS after CRT in patients with locally advanced laryngeal or hypopharyngeal cancer for whom surgical treatment would have required total laryngectomy. We aimed to verify this assumption in the present

retrospective study. We also assessed the interobserver reliability for measuring T-MTV.

## Materials and Methods

**Patients.** Patients with newly diagnosed, technically resectable, stage III/IV laryngeal or hypopharyngeal squamous cell carcinoma requiring total laryngectomy who were definitively treated with CRT between June 2007 and February 2015 at Osaka University Medical School Hospital were eligible. Exclusion criteria were blood glucose levels outside the normal range (80–120 mg/dL) prior to FDG-PET/CT and a >7-week duration between FDG-PET/CT and CRT initiation. In addition to contrast-enhanced CT, FDG-PET/CT has been incorporated into the pretreatment routine work-up since May 2007. The institutional review board approved this study. Written informed consent was waived because of the retrospective nature of the study.

**[18]F-FDG PET/CT and measurement of MTV.** Patients fasted for at least 4 h before they were given i.v. FDG (approximately 3.7 MBq/kg). ^{18}F-FDG PET/CT was carried out using an integrated scanner (Gemini GXL; Philips, Eindhoven, the Netherlands). Whole-body images, generally from the top of the skull to the mid-thigh, were acquired approximately 60 min after i.v. injection of FDG. PET was carried out under the following parameters: 3-D emission scan, 2-min scan per bed position × 11 positions, ordered-subset expectation maximization reconstruction, and 4.0-mm slice thickness per interval. Attenuation correction was done with CT data. Parameters used for CT acquisition were as follows: breath-hold during normal expiration from the lung apex to the lower poles of the kidneys, no i.v. or oral contrast medium, 120 kVp and 50 effective mA, 16 slices, 1.5-mm detector collimation, and 5.0-mm slice thickness with a 4.0-mm interval. Coronal and sagittal CT images were reconstructed using axial thin-section CT images with 1.5-mm slice thickness.

FDG-PET/CT data were transferred into the workstation in the DICOM format. PET/CT parameters were measured from attenuation-corrected PET/CT data using a SUV-based automated contouring program (PETSTAT Viewer Version 2.2 [64 bit]; Adln Research Inc., Tokyo, Japan), which provided an automatically delineated region of interest. The boundary was drawn large enough to incorporate a target lesion in three imaging planes. A series of SUV thresholds were used to define the margin around the tumor, and MTV was automatically calculated. Two head and neck surgeons (J.M., A.H.) with high expertise in interpreting FDG-PET/CT images determined MTV by consensus, with all clinical outcomes blinded. T-MTV and metastatic N-MTV were measured independently. When multiple metastatic nodes were found, MTV of each was measured, the sum of which constituted N-MTV. Whole MTV was defined as the sum of T-MTV and N-MTV.

**CT and image interpretation.** All patients underwent pretreatment CT with a variety of scanners (LightSpeed VCT or Discovery CT 750HD; GE Healthcare, Milwaukee, WI, USA or Aquilion 64 or Aquilion ONE; Toshiba Medical Systems, Otawara, Japan). Axial slices of 2-mm thickness and 2-mm interspace were obtained from the skull base to the upper mediastinum after giving i.v. non-ionic contrast medium. Three-dimensional reconstruction of CT images with proprietary software (AquariusNet Thin Client; TeraRecon Inc., San Mateo, CA, USA) permitted detailed anatomical characterization. A multidisciplinary team interpreted CT images and video records of laryngopharyngeal endoscopic examination to reassess the suitability of total laryngectomy and primary tumor site. T and N classifications were also reassessed according to the AJCC/UICC TNM staging system (seventh edition).[4,8] FDG-PET/CT images were allowed for reference, whereas MTV and clinical outcomes data were blinded.

**Treatment and follow up.** All patients were treated with conventional radiotherapy concurrently with six cycles of weekly docetaxel (10 mg/m$^2$) and cisplatin (20 mg/m$^2$) at a prescribed total dose of 66 Gy in 33 fractions.[9] Response to CRT was assessed by contrast-enhanced CT and inspection by direct laryngoscopy under general anesthesia 10 and 11 weeks after completing CRT, respectively, according to Response Evaluation Criteria in Solid Tumors (RECIST 1.1).[10] The lesser radiographical or clinical response was adopted, as described previously.[9] Subsequently, patients were monitored for recurrence by laryngopharyngeal endoscopy and physical examination every 2 months for 3 years and thereafter every 3 months until death or the final interview. Neck CT scans and plain chest radiography were scheduled every 6 months for 3 years and annually thereafter. CT scans were also planned as necessary. Salvage surgery for persistent and/or recurrent disease was planned when the disease was operable.

**Interobserver reliability of T-MTV measurements.** To evaluate interobserver reliability of T-MTV measurements, five experienced readers (two head and neck surgeons and three radiation oncologists) independently measured T-MTV of 16 patients randomly selected from the study cohort. The same readers also measured T-GTV on CT images that had been used for planning radiotherapy. The CT scanner was equipped with 16 multi-detector rows (BrightSpeed; GE Healthcare). Axial slices of 2.5-mm thickness were obtained from the skull base to the upper mediastinum without contrast enhancement. The obtained CT dataset was transferred to the treatment planning system (XiO; Elekta AB, Stockholm, Sweden). Five readers independently delineated the contour of the primary tumor. All delineation processes were carried out mimicking the actual clinical setting. The readers were allowed to refer to medical charts, video records of laryngopharyngeal endoscopic examination, contrast-enhanced CT images, MRI, and FDG-PET/CT images, whereas the information on T-MTV and clinical outcomes was blinded. Volume of the delineated contour was calculated by the treatment planning system and recorded as T-GTV.

**Statistical analysis.** Time-dependent ROC curve analyses[11] were carried out to obtain the AUC and assess the utility of T-MTV defined by a series of SUV thresholds for predicting disease-specific death. The method developed by DeLong *et al.*[12] was used to examine differences in AUC, and the optimal SUV threshold defining T-MTV was determined. A polynomial curve was fitted to the ROC data and solved for a slope of 1 to identify the T-MTV cut-off value with maximum accuracy for predicting disease-specific death. The cut-off value was applied to dichotomize T-MTV. The 95% CI for AUC and cut-off value were constructed using the bootstrap method with 1000 replications. Events of OS and LFS were defined as any death and laryngectomy or any death, whichever occurred first, respectively. All events were measured from the date of CRT initiation to the date of their occurrence or the date of the last follow-up visit. OS and LFS rates were estimated using the Kaplan–Meier method and compared between groups using the log–rank test. Univariable and multivariable analyses were conducted using a Cox proportional hazards regression model to identify predictors of OS and LFS. Estimated HR and 95% CI were then calculated.

For the multivariable analysis, we generated a T classification model, continuous T-MTV model, and dichotomized T-MTV model as detailed below. Performance of the models was compared using AIC.[13] Differences in AIC followed a $\chi^2$ distribution with 1° of freedom. A logistic regression model was used for univariable and multivariable analyses to identify predictors of incomplete response to CRT, and the estimated OR and 95% CI were calculated. Distant metastasis rates were calculated using the cumulative incidence method, with death without distant metastasis being considered a competing risk, and compared between groups using Gray's test. A Cox proportional hazards regression model was used for univariable and multivariable analyses of distant metastasis. Interobserver reliability for measuring T-MTV and G-MTV was assessed using the ICC and compared using the bootstrap method with 10 000 replications. SpCC was used to evaluate the association between MTV and TNM classification.

All statistical analyses were carried out using SAS for Windows version 9.3 (SAS Institute, Cary, NC, USA), open source statistical software R (http://www.R-progect.org), and JMP version 12 statistical software (IBM Japan, Tokyo, Japan). Two-tailed $P < 0.05$ was considered statistically significant.

## Results

**Patients.** We reviewed clinical records and identified 101 patients with technically resectable stage III/IV laryngeal or hypopharyngeal cancer who underwent definitive CRT. We excluded 14 because total laryngectomy was not necessary, one because of a prolonged interval between FDG-PET/CT and CRT initiation, and one because of a high blood glucose level. Finally, 85 patients were included in the analysis. Patients' baseline characteristics are summarized in Table 1. Local and regional complete response to CRT was observed in 70 and 65 patients, respectively. Twenty-one patients developed distant metastasis. Thirty-six deaths occurred at the time of analysis: 22 as a result of the index cancer, six to another cancer, and eight to other causes. Of the disease-specific deaths, 19 were related to distant metastasis. Ten, three, and one patient underwent total laryngectomy for local residual disease, local recurrent disease, or prelaryngeal nodal recurrence, respectively. With a median follow up of 45 months for surviving patients (range, 12–101 months), 3-year LFS and OS rates for the whole cohort were 56% and 66%, respectively.

**MTV findings.** To determine an optimal SUV threshold delineating MTV, we conducted a time-dependent ROC curve analysis of T-MTV delineated by a series of SUV thresholds (from 2.5 to 5.0, by 0.5 increments), with disease-specific death as the gold standard. AUC of each SUV threshold as a function of time are shown in Figure S1. Although there was no significant difference in AUC among the thresholds at any time point, AUC of the 2.5 SUV threshold tended to be higher than those of other SUV thresholds. Thus, we adopted MTV delineated by a 2.5 SUV threshold for further analyses.

Figure 1(a) depicts T-MTV as a function of T classification. Median T-MTV was 10.0 mL for T2 (range, 3.1–32.0 mL), 12.4 mL for T3 (range, 2.1–42.0 mL), and 28.8 mL for T4a (range, 5.6–88.0 mL). A mild correlation between T-MTV and T classification was observed (SpCC, 0.36; $P = 0.0008$). Likewise, correlations between N-MTV and N classification (SpCC, 0.40; $P = 0.002$) and between whole MTV and TNM stage (SpCC, 0.27; $P = 0.01$) were also mild (Fig. 1b,c).

**Survival.** In the univariable analysis (Table 2), primary tumor site was significantly associated with OS but marginally

**Table 1. Baseline characteristics of patients with technically resectable, stage III/IV laryngeal or hypopharyngeal squamous cell carcinoma requiring total laryngectomy who were definitively treated with CRT**

| Characteristic | No. (%) |
|---|---|
| Age, years | |
|   Median | 66 |
|   Range | 43–79 |
| Gender | |
|   Male | 81 (95) |
|   Female | 4 (5) |
| Primary site | |
|   Hypopharynx | 50 (59) |
|   Larynx | 35 (41) |
| T classification† | |
|   T2 | 19 (22) |
|   T3 | 49 (58) |
|   T4a | 17 (20) |
| N classification† | |
|   N0 | 26 (31) |
|   N1 | 15 (17) |
|   N2a/b/c | 0/29/15 (52) |
| Stage† | |
|   III | 33 (39) |
|   IV | 52 (61) |
| Radiation dose, Gy | |
|   Median | 66 |
|   Range | 64–70 |
|   ≥66 Gy | 84 (99) |
| Chemotherapy cycle | |
|   Median | 6 |
|   Range | 2–6 |
|   6 cycles | 65 (76) |
| Duration between PET and CRT initiation, days | |
|   Median | 25 |
|   Range | 3–49 |

CRT, chemoradiotherapy; PET, positron emission tomography.
†According to the 7th edition of the AJCC/UICC staging system.

with LFS. T-MTV as a continuous variable was a significant determinant of LFS and OS, whereas T classification (T2-T3 versus T4a) was significantly associated with LFS but not OS. N-MTV as a continuous variable was a determinant of neither LFS nor OS, whereas nodal status (negative versus positive) was significantly associated with OS but marginally with LFS. N classification (N0-N1 versus N2) was significantly associated with LFS and OS. TNM stage (stage III versus IV) was significantly associated with both LFS and OS, whereas whole MTV as a continuous variable was a significant determinant of OS but a marginal determinant of LFS.

For the multivariable analysis of LFS, we first constructed two models—T classification model and continuous T-MTV model—with incorporation of primary tumor site and nodal status. As shown in Table 3, after adjustment for primary tumor site and nodal status, patients with T4a disease were at higher risk of laryngectomy or death than patients with T2 or T3 disease (HR, 2.43; 95% CI, 1.20–4.68; $P = 0.01$), and the risk increased by 1.27-fold (95% CI, 1.05–1.50; $P = 0.008$) for every 10-mL increment of T-MTV. According to the AIC, the usefulness of these two models was equivalent ($P = 0.82$). Because T-MTV as a continuous variable was established as an independent predictor of LFS, we stratified T-MTV into
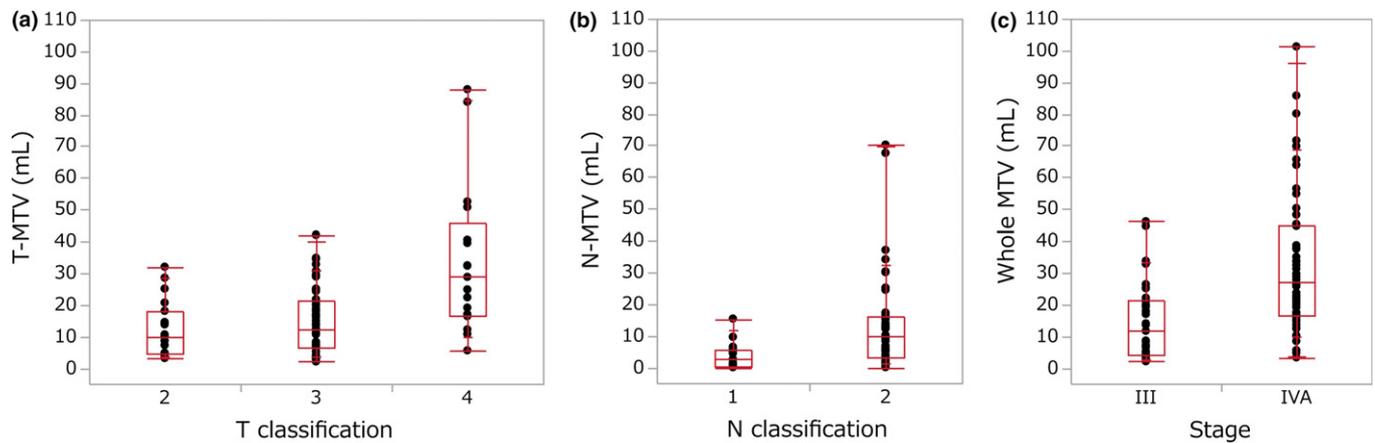
**Fig. 1.** (a) Metabolic tumor volume (MTV) of primary tumor as a function of T classification. (b) Metastatic nodal MTV as a function of N classification. (c) Whole MTV as a function of TNM stage. Each box contains the center 50% of MTV for each classification or stage. Bar within the box indicates the median. Solid lines extending above and below each box indicate the range of MTV. Each dot represents an individual MTV.

**Table 2. Univariable analyses of laryngectomy-free survival and overall survival**

| Variable | No. patients | Laryngectomy-free survival | | | | Overall survival | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | No. events | HR | 95% CI | P-value | No. events | HR | 95% CI | P-value |
| Age (years) | | | | | | | | | |
| ≤65 | 40 | 22 | Reference | | | 17 | Reference | | |
| >65 | 45 | 22 | 0.67 | 0.37–1.23 | 0.20 | 19 | 0.82 | 0.42–1.61 | 0.57 |
| Gender | | | | | | | | | |
| Male | 81 | 43 | Reference | | | 35 | Reference | | |
| Female | 4 | 1 | 0.44 | 0.02–2.03 | 0.35 | 1 | 0.63 | 0.03–2.94 | 0.62 |
| Primary site | | | | | | | | | |
| Larynx | 35 | 13 | Reference | | | 7 | Reference | | |
| Hypopharynx | 50 | 31 | 1.86 | 0.99–3.68 | 0.05 | 29 | 3.55 | 1.64–8.84 | 0.0008 |
| Radiation dose (Gy) | | | | | | | | | |
| ≥66 | 84 | 43 | Reference | | | 35 | Reference | | |
| <66 | 1 | 1 | 16.6 | 0.86–103.1 | 0.06 | 1 | 10.2 | 0.55–55.6 | 0.10 |
| No. chemotherapy cycles | | | | | | | | | |
| 6 cycles | 65 | 35 | Reference | | | 27 | Reference | | |
| <6 cycles | 20 | 9 | 0.80 | 0.36–1.62 | 0.56 | 9 | 1.15 | 0.50–2.40 | 0.71 |
| T classification | | | | | | | | | |
| T2-T3 | 68 | 31 | Reference | | | 26 | Reference | | |
| T4a | 17 | 13 | 2.59 | 1.29–4.91 | 0.009 | 10 | 1.83 | 0.83–3.71 | 0.12 |
| N classification | | | | | | | | | |
| N0-N1 | 41 | 16 | Reference | | | 11 | Reference | | |
| N2 | 44 | 28 | 1.90 | 1.04–3.60 | 0.04 | 25 | 2.62 | 1.32–5.57 | 0.005 |
| Nodal status | | | | | | | | | |
| Negative | 26 | 10 | Reference | | | 5 | Reference | | |
| Positive | 59 | 34 | 1.91 | 0.97–4.10 | 0.06 | 31 | 3.82 | 1.61–11.2 | 0.001 |
| Stage | | | | | | | | | |
| III | 33 | 11 | Reference | | | 8 | Reference | | |
| IV | 52 | 33 | 2.34 | 1.21–4.87 | 0.01 | 28 | 2.60 | 1.23–6.14 | 0.01 |
| T-MTV | | | | | | | | | |
| per 10-mL increment | | | 1.33 | 1.11–1.55 | 0.002 | | 1.43 | 1.17–1.71 | 0.0009 |
| N-MTV | | | | | | | | | |
| per 10-mL increment | | | 0.96 | 0.74–1.16 | 0.72 | | 1.05 | 0.83–1.26 | 0.62 |
| Whole MTV | | | | | | | | | |
| per 10-mL increment | | | 1.14 | 0.99–1.29 | 0.06 | | 1.22 | 1.05–1.40 | 0.009 |

CI, confidence interval; HR, hazard ratio; N-MTV, metabolic tumor volume of metastatic nodes; T-MTV, metabolic tumor volume of primary tumor; whole MTV, whole metabolic tumor volume.

small (≤28.7 mL) and large (>28.7 mL) groups and constructed a dichotomized T-MTV model. Given that 92% of the patients were followed up 2 years after CRT initiation but only 77% at 3 years, we adopted a 2-year cut-off value (28.7 mL; 95% CI, 8.3–40.5) with AUC being 0.81 (95% CI, 0.68–0.96; Fig. S2), although the cut-off values stayed almost equal after

**Table 3.** Multivariable analysis of laryngectomy-free survival

| Variable | No. patients | No. events | T classification model | | | Continuous T-MTV model | | | Dichotomized T-MTV model | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | HR | 95% CI | P-value | HR | 95% CI | P-value | HR | 95% CI | P-value |
| Primary site | | | | | | | | | | | |
| Larynx | 35 | 13 | Reference | | | Reference | | | Reference | | |
| Hypopharynx | 50 | 31 | 1.45 | 0.74–3.00 | 0.27 | 1.34 | 0.67–2.79 | 0.40 | 1.15 | 0.55–2.48 | 0.70 |
| Nodal status | | | | | | | | | | | |
| Negative | 26 | 10 | Reference | | | Reference | | | Reference | | |
| Positive | 59 | 34 | 1.66 | 0.81–3.71 | 0.16 | 1.46 | 0.71–3.25 | 0.30 | 1.34 | 0.62–3.08 | 0.45 |
| T classification | | | | | | | | | | | |
| T2-T3 | 68 | 31 | Reference | | | | | | | | |
| T4a | 17 | 13 | 2.43 | 1.20–4.68 | 0.01 | | | | | | |
| T-MTV | | | | | | | | | | | |
| per 10-mL increment | | | | | | 1.27 | 1.05–1.50 | 0.008 | | | |
| T-MTV (mL) | | | | | | | | | | | |
| ≤28.7 | 68 | 29 | | | | | | | Reference | | |
| >28.7 | 17 | 15 | | | | | | | 4.16 | 1.97–8.70 | 0.0003 |
| | | | AIC: 332.8 | | | AIC: 332.8 | | | AIC: 325.4 | | |

AIC, Akaike's information criterion; CI, confidence interval; HR, hazard ratio; T-MTV, metabolic tumor volume of primary tumor.
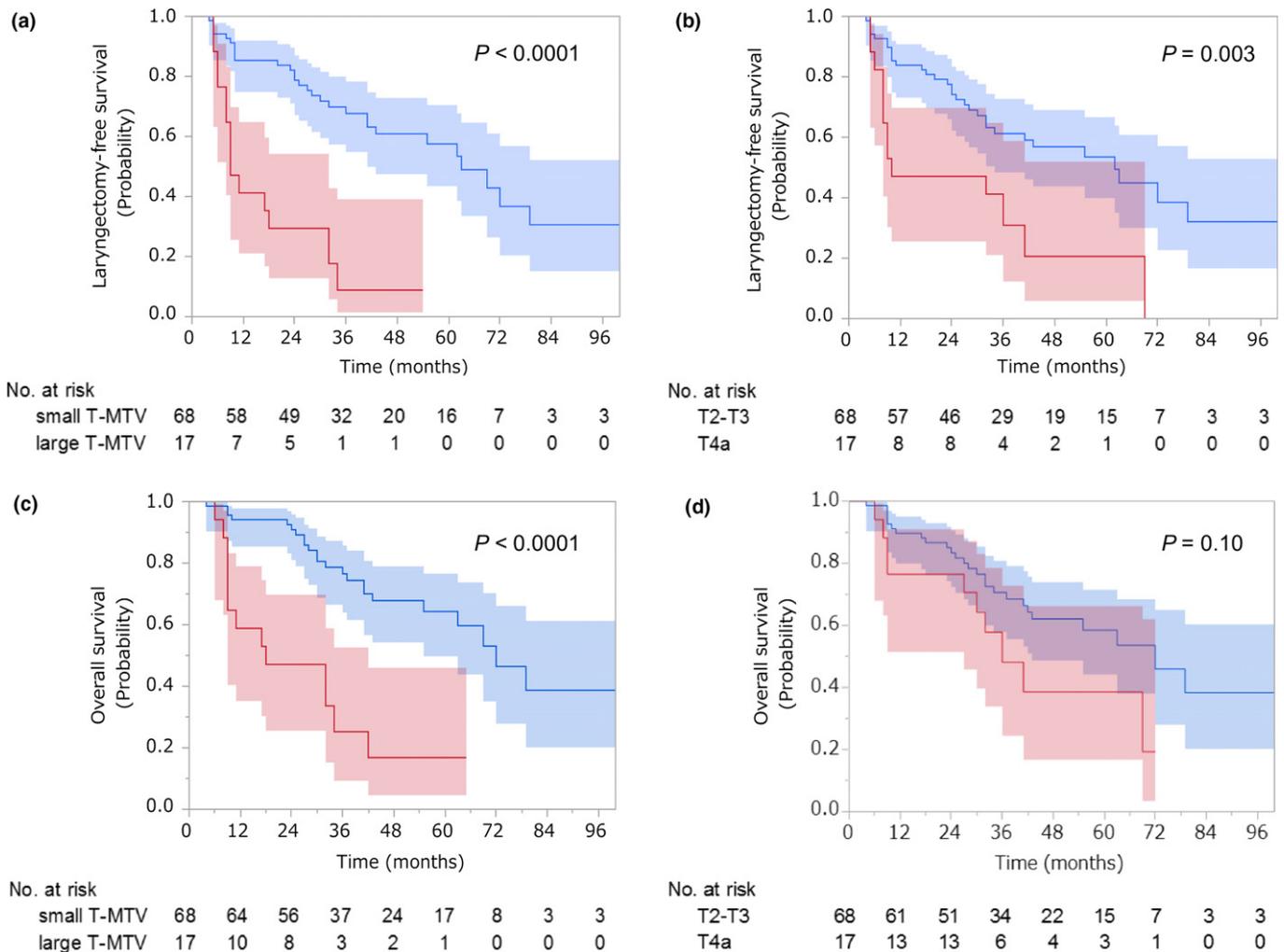


**Fig. 2.** Kaplan–Meier estimates of (a, b) laryngectomy-free survival and (c, d) overall survival with the associated 95% confidence intervals. Patients were stratified by metabolic tumor volume of the primary tumor (T-MTV) in (a) and (c). Blue and red lines represent patients with small (≤28.7 mL) and large (>28.7 mL) T-MTV, respectively. Patients were stratified by T classification in (b) and (d). Blue and red lines represent patients with T2-T3 and T4a disease, respectively.
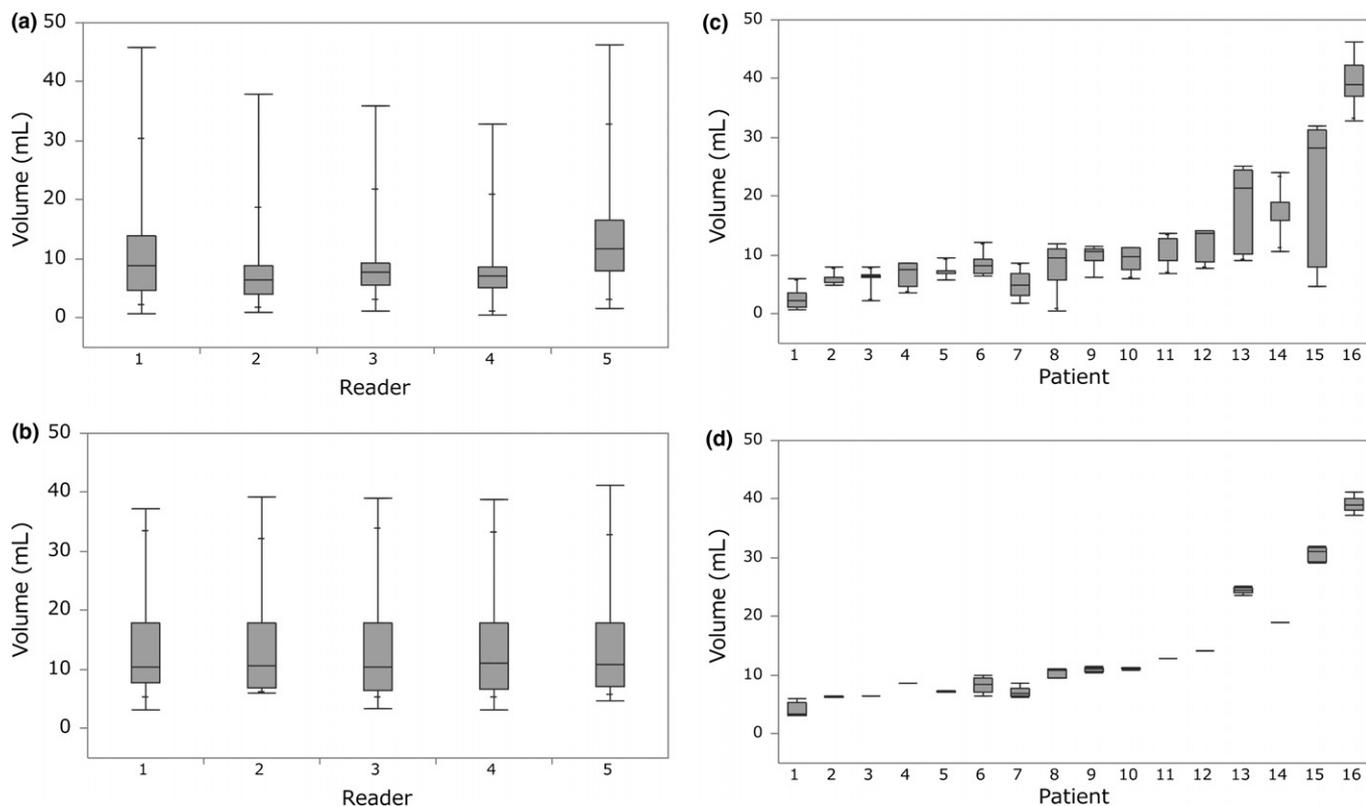
**Fig. 3.** Measurements of (a) gross tumor volume of the primary tumor (T-GTV) and (b) metabolic tumor volume of the primary tumor (T-MTV) provided by five readers for patients are displayed for each reader. Measurements of (c) T-GTV and (d) T-MTV for each patient were provided by five readers. Each box contains the center 50% of the volumes for each reader (a, b) and for each patient (c, d). Bar within the box indicates the median. Solid lines above and below each box indicate the range of volumes.

1 year (Table S1). Risk of laryngectomy or death was higher in patients with large than small T-MTV (HR, 4.16; 95% CI, 1.97–8.70; $P = 0.0003$). Neither primary tumor site nor nodal status was independently associated with LFS in the T-MTV models or T classification model. Dichotomized T-MTV model outperformed the T classification model in predicting LFS (AIC, 325.4 and 332.8, respectively; $P = 0.007$). Kaplan–Meier curves of LFS stratified by T-MTV and T classification are shown in Figure 2(a,b), respectively. Three-year LFS rate was 68% (95% CI, 54–78) in patients with small T-MTV versus 9% (95% CI, 1–30) in patients with large T-MTV ($P < 0.0001$) and 61% (95% CI, 48–72) in patients with T2 or T3 disease versus 31% (95% CI, 10–55) in patients with T4a disease ($P = 0.003$). The difference in LFS between the dichotomized groups was more prominent in the T-MTV model than in the T classification model.

In the multivariable analysis of OS (Table 4), T-MTV remained a significant determinant. Risk of death was increased by 1.28-fold (95% CI, 1.04–1.55; $P = 0.02$) for every 10-mL increment of T-MTV, and patients with large T-MTV were at higher risk than patients with small T-MTV (HR, 3.18; 95% CI, 1.47–6.69; $P = 0.004$). In contrast, T classification was not an independent predictor. Primary tumor site and nodal status were marginally associated with OS in the T-MTV models but were significantly associated with OS in the T classification model. AIC of the T classification model and dichotomized T-MTV model were 265.0 and 259.1, respectively, indicating the advantage of the dichotomized T-MTV model over the T classification model ($P = 0.01$).

Three-year OS rate was 77% (95% CI, 63–86) in patients with small T-MTV versus 25% (95% CI, 7–49) in patients with large T-MTV ($P < 0.0001$; Fig. 2c) and 71% (95% CI, 57–81) in patients with T2 or T3 disease versus 48% (95% CI, 21–71) in patients with T4a disease ($P = 0.10$; Fig. 2d). Difference in OS between the dichotomized groups was significant in the T-MTV model but not in the T classification model. Notably, multivariable analyses results were almost identical when N classification was incorporated instead of nodal status (Table S2).

**Response.** To give insight into the prognostic advantage of T-MTV over T classification, we analyzed their association with local response to CRT and development of distant metastasis. Local response was significantly associated with both T-MTV (irrespective of whether it was a continuous or a dichotomized variable) and T classification, while marginally associated with primary tumor site (Table S3). After adjustment for primary tumor site, T-MTV and T classification remained significant determinants of local response. Risk of incomplete response was increased by 1.52-fold (95% CI, 1.08–2.34; $P = 0.02$) for every 10-mL increment of T-MTV. Large T-MTV disease was at higher risk than small T-MTV disease (OR, 9.79; 95% CI, 2.68–39.8; $P = 0.0005$). Likewise, T4a disease was at higher risk than T2-T3 disease (OR, 6.78; 95% CI, 1.94–24.6; $P = 0.003$). N-MTV as a continuous variable was not a significant determinant of regional response, whereas nodal status was (Table S4). It should be noted that MTV of individual metastatic nodes was also not associated with regional response (data not shown).

**Table 4. Multivariable analysis of overall survival**

| Variable | No. patients | No. events | T classification model | | | Continuous T-MTV model | | | Dichotomized T-MTV model | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | HR | 95% CI | *P*-value | HR | 95% CI | *P*-value | HR | 95% CI | *P*-value |
| Primary site | | | | | | | | | | | |
| Larynx | 35 | 7 | Reference | | | Reference | | | Reference | | |
| Hypopharynx | 50 | 29 | 2.50 | 1.10–6.50 | 0.02 | 2.21 | 0.96–5.77 | 0.06 | 2.08 | 0.86–5.57 | 0.10 |
| Nodal status | | | | | | | | | | | |
| Negative | 26 | 5 | Reference | | | Reference | | | Reference | | |
| Positive | 59 | 31 | 2.84 | 1.12–8.76 | 0.02 | 2.47 | 0.98–7.59 | 0.05 | 2.42 | 0.92–7.61 | 0.07 |
| T classification | | | | | | | | | | | |
| T2-T3 | 68 | 26 | Reference | | | | | | | | |
| T4a | 17 | 10 | 1.88 | 0.84–3.91 | 0.11 | | | | | | |
| T-MTV | | | | | | | | | | | |
| per 10-mL increment | | | | | | 1.28 | 1.04–1.55 | 0.02 | | | |
| T-MTV (mL) | | | | | | | | | | | |
| ≤28.7 | 68 | 23 | | | | | | | Reference | | |
| >28.7 | 17 | 13 | | | | | | | 3.18 | 1.47–6.69 | 0.004 |
| | | | AIC: 265.0 | | | AIC: 262.2 | | | AIC: 259.1 | | |

AIC, Akaike's information criterion; CI, confidence interval; HR, hazard ratio; T-MTV, metabolic tumor volume of primary tumor.

**Distant metastasis.** As shown in Table S5, development of distant metastasis was marginally associated with primary tumor site and T classification and significantly with T-MTV and nodal status, but not with N-MTV as a continuous variable. After adjustment for primary tumor site and nodal status, T-MTV (irrespective of whether a continuous or dichotomized variable) remained a significant determinant of distant metastasis, whereas T classification did not. Risk of distant metastasis was increased by 1.35-fold (95% CI, 1.02–1.73; $P = 0.03$) for every 10-mL increment of T-MTV, and patients with large T-MTV were at higher risk than patients with small T-MTV (HR, 3.08; 95% CI, 1.14–8.07; $P = 0.03$). Three-year cumulative incidence of distant metastasis (Fig. S3) was 21% (95% CI, 12–31) in patients with small T-MTV versus 47% (95% CI, 22–69) in patients with large T-MTV ($P = 0.01$).

**Interobserver reliability.** Eighty measurements for T-MTV and T-GTV were carried out (16 patients, five readers). ICC for T-MTV and T-GTV were 0.99 (95% CI, 0.99–1.00) and 0.85 (95% CI, 0.68–0.94), respectively. Difference between the two was statistically significant ($P < 0.0001$), indicating higher interobserver reliability when measuring T-MTV than T-GTV. Distributions of T-MTV and T-GTV are shown in Figure 3(a,b), respectively, for each reader and in Figure 3(c,d), respectively, for each patient. Overlap across readers and narrow distribution among patients were more evident for T-MTV than for T-GTV.

## Discussion

We demonstrated that pretreatment T-MTV served as an independent predictor of LFS and OS in patients with locally advanced laryngeal or hypopharyngeal cancer requiring total laryngectomy when treated with CRT. We also demonstrated that pretreatment T-MTV outperformed T classification in stratifying patients into low- and high-risk groups with respect to LFS and OS.

Both T-MTV and T classification were independently associated with local response to CRT, whereas only T-MTV was a significant determinant of distant metastasis. Given that the vast majority of disease-specific deaths were related to distant metastasis, distant metastasis was a key determinant of survival. Difference in the association with distant metastasis

between T-MTV and T classification was most likely responsible for the prognostic advantage of T-MTV over T classification. The prognostic significance also differed between T-MTV and N-MTV. T-MTV was predictive of LFS and OS, whereas N-MTV was predictive of neither. T-MTV was associated with the development of distant metastasis whereas N-MTV was not, although the reason remains unclear. This plausibly underlies the prognostic difference between T-MTV and N-MTV. Whole MTV was predictive of OS, which was probably attributed to the strong association of T-MTV with OS.

Notably, nodal status was significantly associated with OS, whereas N-MTV as a continuous variable was not. Moreover, the optimal N-MTV cut-off determined by time-dependent ROC curve analysis was as small as 0.91 mL at 1 year after CRT initiation and thereafter (data not shown). Given that most patients with node-positive disease showed N-MTV of >0.91 mL, the presence or absence, rather than the volumetric size, of metastatic nodes is likely important for predicting survival of patients undergoing CRT in the setting of larynx preservation. Although this assumption is supported by the finding that not N-MTV as a continuous variable but nodal status was associated with distant metastasis, it needs to be validated in a larger cohort that includes N3 disease.

The finding that T-MTV was superior to T classification in predicting LFS and OS prompts us to assume that T-MTV, instead of T classification, could serve to individualize treatment for patients requiring total laryngectomy, such as CRT aiming at larynx preservation for patients with small T-MTV and upfront total laryngectomy for patients with large T-MTV. However, because our findings are based on retrospective data, such an assumption requires prospective validation. Additionally, how T-MTV is associated with survival of patients undergoing upfront total laryngectomy remains unclear. To address this problem, we are currently analyzing a different cohort of patients who underwent upfront total laryngectomy.

It could be argued that T-GTV defined on CT images is an alternative to T-MTV. Previous efforts demonstrated that when patients with locally advanced laryngeal or hypopharyngeal cancer were treated with radiotherapy alone or CRT, those with small T-GTV survived better than those with large T-GTV.[14,15] However, non-negligible observer variation is present when measuring GTV.[6,16]

This is because estimation of tumor shape on CT images varies among observers.[17] Given the excellent interobserver reliability of MTV compared with GTV in our cohort, we favor MTV, which can be calculated quickly using semiautomatic contouring tools.

Our study has some limitations. First, it was retrospective. Second, the chemotherapy in CRT was weekly low-dose cisplatin/docetaxel, although tri-weekly high-dose cisplatin is the standard.[18] Given that a phase II study evaluating the efficacy of our CRT regimen was successful,[9] we assume that the cut-off value of T-MTV obtained in this study will serve to dichotomize the same line of patients treated with the standard CRT. This assumption is currently being validated.

In conclusion, pretreatment T-MTV would be a more useful predictor of LFS and OS than T classification when patients with laryngeal or hypopharyngeal cancer requiring total laryngectomy are treated with the larynx preservation approach, although the usefulness of T-MTV needs further validation. Given the excellent interobserver reliability of T-MTV measurements, T-MTV could help guide a novel clinical trial design in the setting of larynx preservation.

## Disclosure Statement

Authors declare no conflicts of interest for this article.

## Abbreviations

| | |
|---|---|
| $^{18}$F-FDG PET | positron emission tomography with $^{18}$F-fluorodeoxyglucose |
| AIC | Akaike's information criterion |
| AJCC | American Joint Committee on Cancer |
| AUC | area under the curve |
| CI | confidence interval |
| CRT | chemoradiotherapy |
| CT | computed tomography |
| DICOM | Digital Imaging and Communications in Medicine |
| HR | hazard ratio |
| ICC | intraclass correlation coefficient |
| LFS | laryngectomy-free survival |
| MTV | metabolic tumor volume |
| N-MTV | MTV of metastatic nodes |
| OR | odds ratio |
| OS | overall survival |
| ROC | receiver operating characteristic |
| SpCC | Spearman's correlation coefficient |
| SUV | standardized uptake value |
| T-GTV | gross tumor volume of primary tumor |
| T-MTV | MTV of primary tumor |

## References

1 Lefebvre JL, Ang KK. Larynx Preservation Consensus Panel. Larynx preservation clinical trial design: key issues and recommendations–a consensus panel summary. *Head Neck* 2009; **31**: 429–41.
2 National Comprehensive Cancer Network. Guidelines: Head and Neck Cancers, 2017. [Cited 01 May 2017.] Available from URL: http://www.nccn.org/professionals/physician_gls/pdf/head-and-neck.pdf
3 Stokes WA, Jones BL, Bhatia S *et al.* A comparison of overall survival for patients with T4 larynx cancer treated with surgical versus organ-preservation approaches: A National Cancer Data Base analysis. *Cancer* 2017; **123**: 600–8.
4 Edge SB, Byrd DR, Compton CC *et al.*, eds. *AJCC Cancer Staging Handbook*, 7th edn. New York, NY: Springer, 2010.
5 Beitler JJ, Muller S, Grist WJ *et al.* Prognostic accuracy of computed tomography findings for patients with laryngeal cancer undergoing laryngectomy. *J Clin Oncol* 2010; **28**: 2318–22.
6 Hoorweg JJ, Kruijt RH, Heijboer RJ, Eijkemans MJ, Kerrebijn JD. Reliability of interpretation of CT examination of the larynx in patients with glottic laryngeal carcinoma. *Otolaryngol Head Neck Surg* 2006; **135**: 129–34.
7 Hanamoto A, Tatsumi M, Takenaka Y *et al.* Volumetric PET/CT parameters predict local response of head and neck squamous cell carcinoma to chemoradiotherapy. *Cancer Med* 2014; **3**: 1368–76.
8 Sobin L, Gospodarowicz M, Wittekind C, eds. *UICC TNM Classification of Malignant Tumours*, 7th edn. Chichester, UK: Wiley-Blackwell, 2010.
9 Inohara H, Takenaka Y, Yoshii T *et al.* Phase 2 study of docetaxel, cisplatin, and concurrent radiation for technically resectable stage III-IV squamous cell carcinoma of the head and neck. *Int J Radiat Oncol Biol Phys* 2015; **91**: 934–41.

10 Eisenhauer EA, Therasse P, Bogaerts J *et al.* New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009; **45**: 228–47.
11 Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat Med* 2013; **32**: 5381–97.
12 DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988; **44**: 837–84.
13 Akaike H. A new look at the statistical model identification. *IEEE Trans Autom Control* 1974; **19**: 16–23.
14 Chen SW, Yang SN, Liang JA, Lin FJ, Tsai MH. Prognostic impact of tumor volume in patients with stage III-IVA hypopharyngeal cancer without bulky lymph nodes treated with definitive concurrent chemoradiotherapy. *Head Neck* 2009; **31**: 709–16.
15 Mancuso AA, Mukherji SK, Schmalfuss I *et al.* Preradiotherapy computed tomography as a predictor of local control in supraglottic carcinoma. *J Clin Oncol* 1999; **17**: 631–7.
16 Mukherji SK, Toledano AY, Beldon C *et al.* Interobserver reliability of computed tomography-derived primary tumor volume measurement in patients with supraglottic carcinoma. *Cancer* 2005; **103**: 2616–22.
17 Cooper JS, Mukherji SK, Toledano AY *et al.* An evaluation of the variability of tumor-shape definition derived by experienced observers from CT images of supraglottic carcinomas (ACRIN protocol 6658). *Int J Radiat Oncol Biol Phys* 2007; **67**: 972–5.
18 Pignon JP, le Maitre A, Maillard E *et al.* Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): an update on 93 randomized trials and 17,346 patients. *Radiother Oncol* 2009; **92**: 4–14.

## Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

**Fig. S1.** AUC of a series of SUV thresholds (from 2.5 to 5.0, by 0.5 increments) delineating T-MTV were determined by time-dependent receiver operating characteristic curve analysis of T-MTV for prediction of disease-specific death for the first 5 years after the start of chemoradiotherapy.

**Fig. S2.** Time-dependent ROC curve analysis for predicting disease-specific death according to T-MTV 2 years after the start of chemoradiotherapy.

**Fig. S3.** Cumulative incidence of distant metastasis among patients stratified by T-MTV, with death without distant metastasis being a competing risk.

**Table S1.** Time-dependent ROC analysis of disease-specific death.

**Table S2.** Multivariable analysis of laryngectomy-free survival and overall survival.

**Table S3.** Univariable and multivariable analysis of local response.

**Table S4.** Univariable analysis of regional response.

**Table S5.** Univariable and multivariable analysis of distant metastasis.