PET-based Risk Stratification in Primary Mediastinal B Cell Lymphoma: A Comparative Analysis of different segmentation methods in the IELSG37 trial patient cohort.

Luca Ceriani[1,2], Lisa Milan[1], Maria Cristina Pirosa[3], Maurizio Martelli[4], Teresa Ruberto[1], Luciano Cascione[2], Peter W.M. Johnson[5], Andrew J. Davies[5], Giovannino Ciccone[6] and Emanuele Zucca[2,3].

[1] Imaging Institute of Southern Switzerland - EOC, Lugano, Switzerland

[2] Institute of Oncology Research, Faculty of biomedical sciences, Università della Svizzera Italiana, Bellinzona, Switzerland

[3] Oncology Institute of Southern Switzerland - EOC, Bellinzona, Switzerland

[4] Hematology Department of Translational and Precision Medicine, Sapienza University, Roma, Italy

[5] Cancer Research UK Centre, University of Southampton, Southampton, UK

[6] Clinical Epidemiology Unit, AOU Città della Salute e della Scienza di Torino and CPO Piemonte, Turin, Italy

Word counts = 2600

Corresponding author: Luca Ceriani, Ospedale Regionale Lugano, Via Tessere 46, CH-6900 Lugano.

E-mail:luca.ceriani@eoc.ch.Telephone +4191 811 6446 (reprints are not available)

Financial Support: This study was funded in part by Swiss Cancer Research foundation (KLS-5406-08-2021).

Running title: PET metrics estimation in PMBCL

ABSTRACT

Rationale: Standardizing tumor measurement on ^{18}F-FDG-PET is crucial for the routine clinical use of powerful PET-derived lymphoma prognostic factors like metabolic tumor volume (MTV) and tumor lesion glycolysis (TLG). The recent proposal of SUV≥4 as a new reference segmentation threshold for most aggressive lymphomas may homogenize volume-based metrics and facilitate their clinical application. **Methods:** This study compared MTV and TLG in primary mediastinal B-cell lymphoma (PMBCL) patients estimated using SUV≥4 and the current threshold at 25% of SUVmax. Baseline PET-metrics were evaluated in 501 PMBCL patients from the IELSG37 trial. **Results:** Median MTV and TLG estimated with the 25%SUVmax threshold were significantly lower than those obtained with the new reference threshold, however an extremely high correlation was observed between the methods for both MTV (r=0.95) and TLG (r=0.99), resulting in superimposable prognostic power. **Conclusions:** These findings support the routine use of the threshold at SUV≥4 for volumetric measurements in PMBCL.

KEY WORDS

Positron emission tomography

Primary mediastinal B-cell lymphoma

Metabolic tumor volume

INTRODUCTION

The functional parameters metabolic tumor volume (MTV) and total lesion glycolysis (TLG) derived from positron emission tomography/computed tomography with ^{18}F-fluorodeoxyglucose (^{18}F-FDG PET/CT) are promising prognostic factors in different malignant lymphoma subtypes (*1,2*). In practice, these parameters are often categorized identifying optimal cutoffs to distinguish high-risk from low-risk patients. However, these cut-offs depend largely on population characteristics and the method used to segment tumor lesions on PET images. Different threshold values have been applied, including absolute (SUV = 2.5 or 4), adaptive (1.5 times the mean SUV of the liver ±2 standard deviations), and relative (41% of SUVmax) to estimate MTV (*3-8*). Although the different methods produced different MTV cutoffs, their prognostic impact was comparable (*9*). Consequently, a universally applicable cutoff for MTV to classify patient prognosis remains elusive. Standardization of MTV measurements is essential for its routine use as a prognostic marker (*10*). In primary mediastinal B-cell lymphoma (PMBCL), unlike other subtypes, a threshold of 25% of SUVmax has been identified as optimal for measuring large masses, with TLG estimated emerging as the best prognostic marker (*11*). Recent emphasis has been placed on using SUV≥4 as a threshold for MTV measurements in FDG-avid lymphomas due to reduced observer interaction and lower inter-operator variability (*12*). This fixed threshold has been proposed as a reference method for tumor lesion segmentation and MTV estimation in follicular lymphoma, diffuse large B-cell lymphoma, and Hodgkin lymphoma (*13*). However, its application in PMBCL remains untested. This study aims to compare MTV and TLG values estimated using SUV≥4 and 25% of SUVmax thresholds, evaluating their prognostic properties in a large cohort of PMBCL patients enrolled in the IELSG37 trial(*14*).

METHODS

Baseline ^{18}F-FDG PET scans were evaluated in 501 PMBCL patients enrolled between 2012 and 2019 in the IELSG37 study (NCT01599559) (14). A cohort of 103 PMBCL patients enrolled between 2007 and 2010 in the previous IELSG26 study (NCT00944567) (11) was used for validation. Lesion contouring was performed using a dedicated software (MM-Oncology Syngo.via VB60A, Siemens) with two thresholds (SUV≥4 and 25% of SUVmax). MTV was estimated for each threshold and TLG was calculated as the product of mean SUV and MTV. Data were presented as median and interquartile range (IQR). PET metrics were compared using Wilcoxon's signed-rank test, linear regression and Pearson correlation. Optimal cut-points were determined by ROC analysis using the Youden method with Fluss adjustment for continuous variables (*15*). Progression-free survival (PFS) and overall survival (OS) were estimated using the Kaplan-Meier or life-table methods, with differences analyzed by the log-rank test. Multivariable analysis was performed using Cox proportional hazard models. Predictive accuracy of the PET parameters was compared using the Gonen and Heller concordance probability index (*16*) and the Akaike information criterion (*17*). Statistical analyses were conducted using STATA 18 (StataCorp). The relevant ethics committees approved the study, and all patients gave written consent.

RESULTS

SUVmax ranged between 4.54 to 48.99 (median, 21.48 [IQR, 17.59-26.77]). The actual SUV value used for segmentation with the 25% of SUVmax threshold was higher than 4 in over 75% of patients (median, 5.37 [IQR, 4.40-6.70]). Consequently, the median MTV estimated with the 25% of SUVmax threshold was lower than that obtained with the SUV≥4 threshold (331 ml [IQR, 201-530] vs. 396 ml [IQR, 229-635]; *P* <0.0001) [Figure 1A]. TLG showed a similar difference (3387 [IQR, 1894-6070] vs. 3750 [IQR, 1953-6425] *P* <0.0001) [Figure 1B]. Despite these differences, MTV values calculated by both methods were highly correlated (r=0.95; *P* <0.001) [Figure 1C], with an even stronger correlation for TLG (r=0.99; *P*<0.001) [Figure 1D].

In univariate Cox regression analysis, higher baseline MTV and TLG levels as continuous variables were significantly associated with an increased risk of progression or death. As expected, this association persisted across both PFS and OS when the segmentation threshold was changed from 25% of SUVmax to SUV≥4 (Table 1).

ROC analysis identified the optimal cutoff points for both MTV and TLG using each segmentation method (Table 2). These methods demonstrated similar risk-stratification capabilities.

High TLG values consistently indicated poorer PFS, regardless of the threshold used. Risk class changes occurred in 27 patients (5%), none with extranodal involvement. Seventeen shifted from high to low risk (median SUVmax, 19.88 [IQR, 14.38-23.08]), and 10 from low to high risk (median SUVmax, 30.79 [IQR, 25.99-32.09]). Only 2 patients, who moved from high to low risk when switching from 25% of SUVmax to SUV≥4, experienced disease progression, while the remaining 25 had no events.

Similarly, PFS was significantly shorter in patients with high MTV (Figure 2), despite risk category changes in 49 patients (10%). Risk decreased in 26 (median SUVmax, 13.22 [IQR, 10.77-16.45]) and increased in 23 (median SUVmax, 30.52 [IQR, 27.49-33.05]). Only 3 patients with progressive disease shifted from high to low risk, with no events in the other 46.

The univariate analysis of OS yielded similar findings (Table 2, Figure 3). No consistent superiority was found between the two thresholding methods for predicting PFS or OS (Table 2).

The ability of MTV and TLG using the SUV≥4 method to discriminate patients with significantly different PFS and OS was confirmed in the 103-patient IELSG26 cohort, previously used to demonstrate the superiority of the 25% SUVmax threshold in PMBCL [Supplemental Figure 1].

DISCUSSION

Despite the emergence of baseline volume-based PET metrics as important prognostic factors in lymphomas, the lack of standardized measurement procedures has limited their use in routine practice and clinical trials (*5,9*). The commonly used thresholds for segmentation in lymphomas and solid cancers (*8*) are suboptimal for PMBCL, a rare non-Hodgkin lymphoma subtype typically characterized by a single bulky mass in the anterior mediastinum (*14*). In a prior study, we found that a 25% of SUVmax threshold is more appropriate for accurately segmenting these large masses, with TLG emerging as the best baseline predictor of outcomes in PMBCL (*11*).

A recent benchmark study reached a consensus among international experts, recommending an SUV≥4.0 threshold as the standard for tumor segmentation and MTV estimation in FDG-avid lymphomas across various subtypes (*13*). This method was favored for its ease of implementation, reproducibility, and minimal reader variability (*13*). The semi-automatic segmentation approach used ensures high inter-operator reproducibility, including only lesions with SUV≥4 and a volume of ≥3 mL, with minimal manual adjustments for physiological uptake (*13*). However, this study did not include PMBCL patients.

In this study, using the largest cohort of prospectively enrolled PMBCL patients (*14*), we observed strong correlations between MTV and TLG values obtained using both methods. Notably, the SUV≥4 threshold effectively segmented large mediastinal tumors without reducing the predictive accuracy of volumetric PET metrics compared to the 25% of SUVmax method.

CONCLUSION

Our findings suggest that the SUV≥4 threshold is a reliable contouring method applicable to all FDG-avid lymphoma subtypes, including PMBCL. Adopting this threshold as a standard could enhance the consistency of risk stratification strategies in future clinical trial designs.

FIGURE Legends

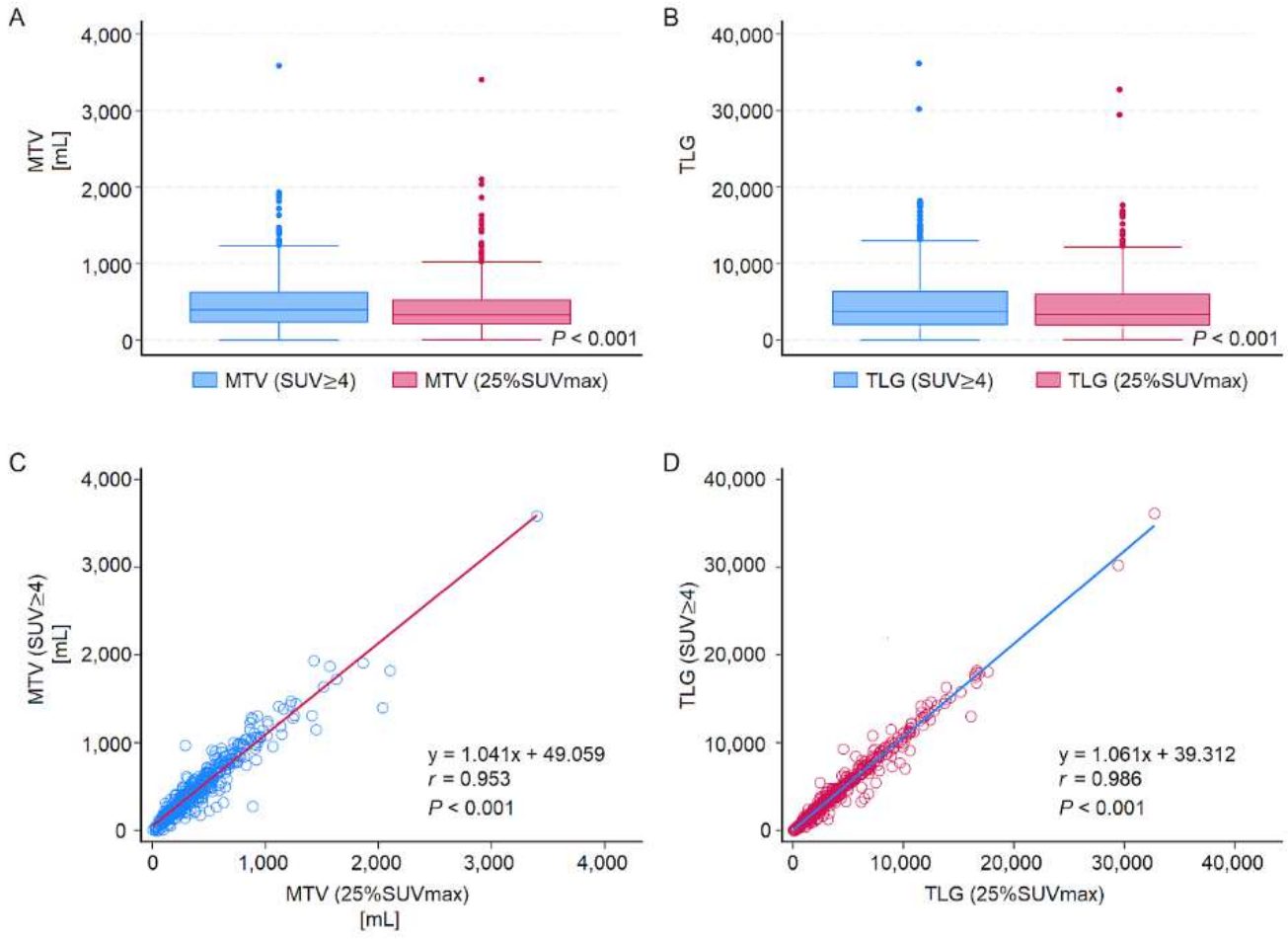

Figure 1. Comparison of MTV (A) and TLG (B) estimated using two segmentation thresholds: 25% of SUVmax and SUV≥4. Pearson correlation plots for MTV (C) and TLG (D) values calculated with each segmentation method are also shown.

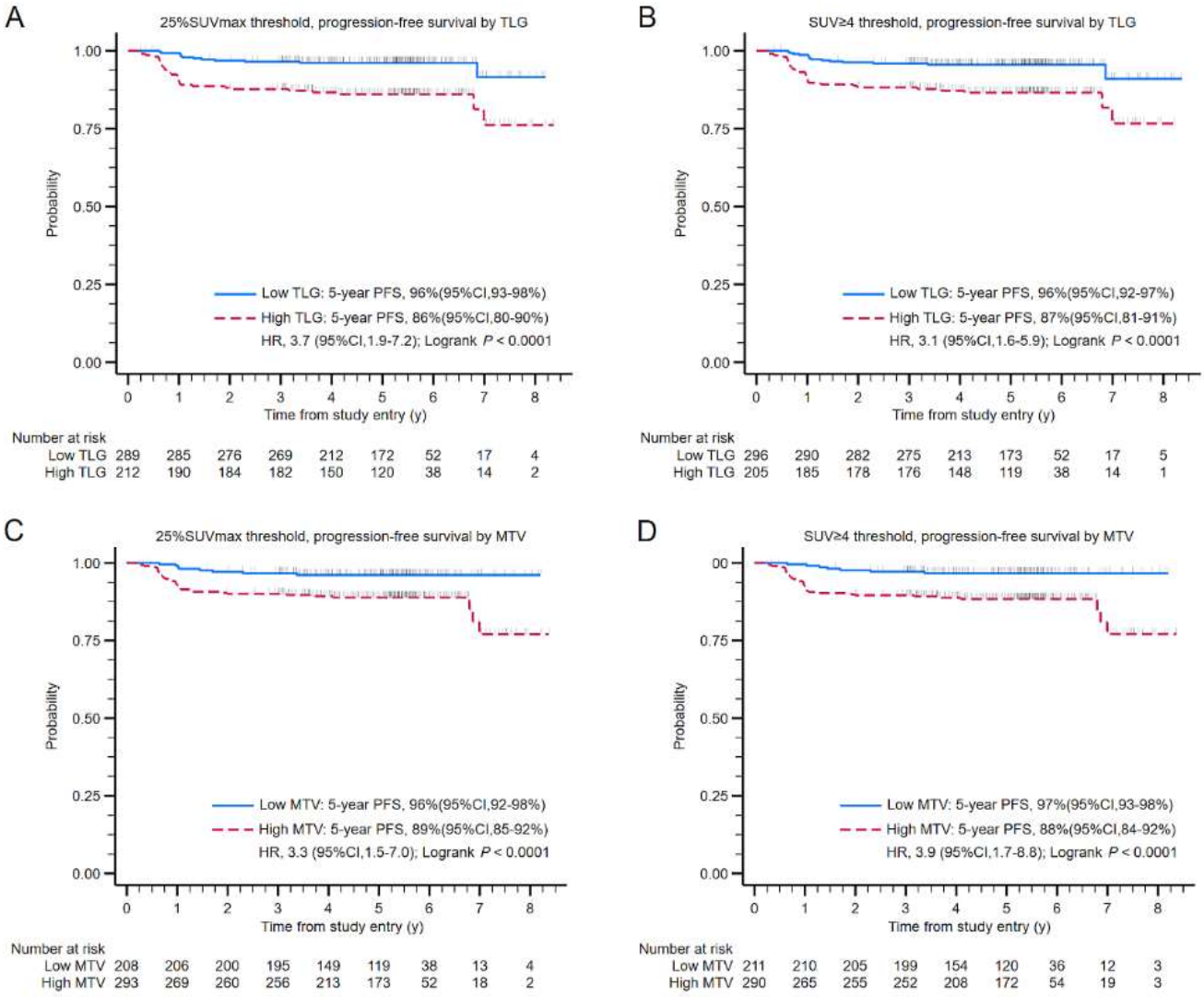

Figure 2. Kaplan-Meier estimates of progression-free survival according to dichotomized TLG (A and B) and MTV (C and D) estimated using either the 25% of SUVmax (A and C) or the SUV≥4 (B and D) thresholding method in the IELS37 study population (N=501).

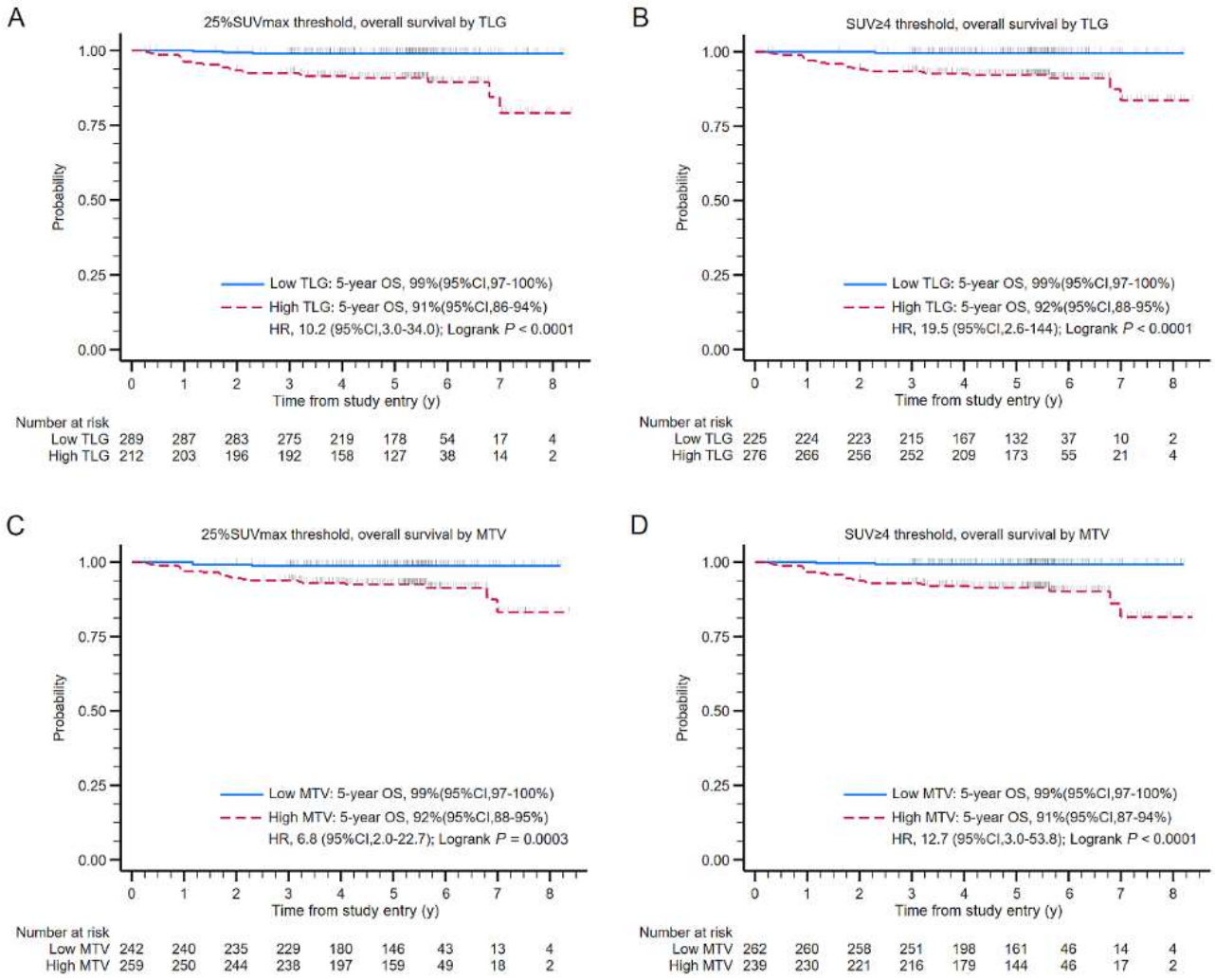

Figure 3. Kaplan-Meier estimates of overall survival according to dichotomized TLG (A and B) and MTV (C and D) estimated using either the 25% of SUVmax (A and C) or the SUV≥4 (B and D) thresholding method in the IELS37 study population (N=501).

TABLES

Table 1. Univariate analysis (Cox regression) of the impact of increments of 10^2ml for baseline MTV and increments of 10^3 for TLG as continuous variables on PFS (43 events) and OS (25 events) in the IELSG37 study cohort (N=501)

	Thresholding Method	PET parameter	HR (95%CI)	P-value
Progression-free survival	25%SUVmax	TLG	1.078 (1.025-1.134)	0.003
	25%SUVmax	MTV	1.083 (1.027-1.142)	0.003
	SUV≥4	TLG	1.075 (1.027-1.126)	0.002
	SUV≥4	MTV	1.089 (1.038-1.142)	0.000
Overall survival	25%SUVmax	TLG	1.112 (1.054-1.173)	0.000
	25%SUVmax	MTV	1.113 (1.055-1.175)	0.000
	SUV≥4	TLG	1.105 (1.052-1.160)	0.000
	SUV≥4	MTV	1.116 (1.062-1.172)	0.000

HR, Hazard ratio; CI, confidence interval,

TABLE 2. Analysis of PFS and OS in the IELSG37 cohort: Comparison of predictive accuracy of MTV and TLG estimated using 25%SUVmax and SUV≥4 thresholds.

	Thresholding Method	Dichotomized Variable (N)	Cutpoint	AUC (95%CI)	Percent Surviving at 5 Years (95%CI)	HR (95%CI)	P-Value	Gonen & Heller's *K* (SE)	Akaike's Information Criterion*
Progression-free survival	25%SUVmax	Low TLG (289) High TLG (212)	3865.91	0.648 (0.604-0.690)	96.1 (93.1-97.8) 86.0 (80.5-90.1)	3.7 (1.9-7.2)	<0.0001	0.6405 (0.0279)	499.16
	25%SUVmax	Low MTV (208) High MTV (293)	282.95	0.639 (0.595-0.681)	96.1 (92.3-98.0) 88.8 (84.5-92.0)	3.3 (1.5-7.0)	0.0014	0.6290 (0.0343)	504.78
	SUV≥4	Low TLG (296) High TLG (205)	4227.24	0.649 (0.606-0.691)	95.5 (92.4-97.4) 86.5 (80.9-90.6)	3.3 (1.6-5.9)	0.0002	0.6243 (0.0291)	502.88
	SUV≥4	Low MTV (211) High MTV (290)	337.15	0.656 (0.613-0.698)	96.6 (93.0-98.4) 88.3 (84.0-91.6)	3.9 (1.7-8.8)	0.0004	0.6450 (0.0327)	501.73
Overall survival	25%SUVmax	Low TLG (289) High TLG (212)	3865.91	0.753 (0.713-0.790)	98.9 (96.8-99.7) 90.7 (85.8-94.0)	10.2 (3.0-34)	<0.0001	0.7008 (0.0247)	274.68
	25%SUVmax	Low MTV (242) High MTV (259)	318.3	0.726 (0.685-0.765)	98.7 (96.1-99.6) 92.4 (88.3-95.1)	6.8 (2.0-23)	0.0003	0.6861 (0.0344)	282.97
	SUV≥4	Low TLG (225) High TLG (276)	3355.14	0.751 (0.710-0.788)	99.5 (96.8-99.9) 92.1 (88.2-94.8)	19.5 (2.7-144)	<0.0001	0.7238 (0.0235)	276.15
	SUV≥4	Low MTV (262) High MTV (239)	414.75	0.752 (0.712-0.825)	99.2 (96.9-99.8) 91.3 (86.9-94.3)	12.7 (3.0-54)	<0.0001	0.7134 (0.0250)	275.08

HR, Hazard ratio; CI, confidence interval; SE, standard error; AUC, area under the receiver operating characteristic (ROC) curve.

*The Gonen and Heller's concordance probability estimator (*K*) ranges from 0.5 to 1, and higher values indicate more accurate discrimination; the Akaike information criterion (AIC) compares the relative quality of a set of statistical models (the optimal model is the one with minimum AIC).

DISCLOSURE

The authors have no conflicts of interest to disclose. This work was partly funded by the Swiss Cancer Research Foundation (KLS-5406-08-2021). The IELSG is supported by the Swiss State Secretariat for Education, Research and Innovation and the Swiss Cancer Research Foundation.

KEY POINTS

QUESTION: Is it possible to use the SUV≥4 threshold – which was recently proposed as a new standard for diffuse large B-cell lymphomas, follicular lymphomas, and Hodgkin lymphoma – also in PET/CT segmentation of primary mediastinal B cell lymphomas, which until now have been specifically segmented with a 25% of SUVmax threshold?

PERTINENT FINDINGS: Functional PET parameters generated using two segmentation methods (SUV≥4 vs. 25% of SUVmax) in a large prospective clinical trial cohort demonstrated similar prognostic capabilities.

IMPLICATIONS FOR PATIENT CARE: Our findings suggest that the SUV≥4 threshold is a reliable contouring method applicable to all FDG-avid lymphoma subtypes, including primary mediastinal lymphomas. Adopting this threshold procedure may improve the consistency of risk stratification in lymphoma clinical trials.

References

1. Meignan M, Cottereau AS, Specht L, Mikhaeel NG. Total tumor burden in lymphoma - an evolving strong prognostic parameter. *Br J Radiol.* 2021;94:20210448.

2. Guo B, Tan X, Ke Q, Cen H. Prognostic value of baseline metabolic tumor volume and total lesion glycolysis in patients with lymphoma: A meta-analysis. *PLoS One.* 2019;14:e0210224.

3. Kostakoglu L, Mattiello F, Martelli M, et al. Total metabolic tumor volume as a survival predictor for patients with diffuse large B-cell lymphoma in the GOYA study. *Haematologica.* 2022;107:1633-1642.

4. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors. *J Nucl Med.* 2009;50 Suppl 1:122s-150s.

5. Ceriani L, Milan L, Johnson PWM, et al. Baseline PET features to predict prognosis in primary mediastinal B cell lymphoma: a comparative analysis of different methods for measuring baseline metabolic tumour volume. *Eur J Nucl Med Mol Imaging.* 2019;46:1334-1344.

6. Ilyas H, Mikhaeel NG, Dunn JT, et al. Defining the optimal method for measuring baseline metabolic tumour volume in diffuse large B cell lymphoma. *Eur J Nucl Med Mol Imaging.* 2018;45:1142-1154.

7. Eude F, Toledano MN, Vera P, Tilly H, Mihailescu SD, Becker S. Reproducibility of Baseline Tumour Metabolic Volume Measurements in Diffuse Large B-Cell LymphomA: Is There a Superior Method? *Metabolites.* 2021;11.

8. Boellaard R, Delgado-Bolton R, Oyen WJ, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *Eur J Nucl Med Mol Imaging.* 2015;42:328-354.

9. Barrington SF, Meignan M. Time to Prepare for Risk Adaptation in Lymphoma by Standardizing Measurement of Metabolic Tumor Burden. *J Nucl Med.* 2019;60:1096-1102.

10. Barrington SF, Cottereau AS, Zijlstra JM. Is (18)F-FDG Metabolic Tumor Volume in Lymphoma Really Happening? *J Nucl Med.* 2024;65:510-511.

11. Ceriani L, Martelli M, Zinzani PL, et al. Utility of baseline 18FDG-PET/CT functional parameters in defining prognosis of primary mediastinal (thymic) large B-cell lymphoma. *Blood.* 2015;126:950-956.

12. Barrington SF, Zwezerijnen B, de Vet HCW, et al. Automated Segmentation of Baseline Metabolic Total Tumor Burden in Diffuse Large B-Cell Lymphoma: Which Method Is Most Successful? A Study on Behalf of the PETRA Consortium. *J Nucl Med.* 2021;62:332-337.

13. Boellaard R, Buvat I, Nioche C, et al. International Benchmark for Total Metabolic Tumor Volume Measurement in Baseline (18)F-FDG PET/CT of Lymphoma Patients: A Milestone Toward Clinical Implementation. *J Nucl Med.* 2024;65:1343-1348.

14. Martelli M, Ceriani L, Ciccone G, et al. Omission of Radiotherapy in Primary Mediastinal B-Cell Lymphoma: IELSG37 Trial Results. *J Clin Oncol.* 2024;42:4071-4083.

15. Fluss R, Faraggi D, Reiser B. Estimation of the Youden Index and its associated cutoff point. *Biom J.* 2005;47:458-472.

16. Gönen M, Heller G. Concordance probability and discriminatory power in proportional hazards regression. *Biometrika.* 2005;92:965-970.

17. Posada D, Buckley TR. Model Selection and Model Averaging in Phylogenetics: Advantages of Akaike Information Criterion and Bayesian Approaches Over Likelihood Ratio Tests. *Systematic Biology.* 2004;53:793-808.

GRAPHICAL ABSTRACT

Strong Correlation and Comparable Prognostic Value of MTV Using 25% SUVmax and SUV≥4 Thresholds for Lesion Segmentation in PET Scans of Primary Mediastinal B-cell Lymphoma