ORIGINAL ARTICLE

Baseline PET features to predict prognosis in primary mediastinal B cell lymphoma: a comparative analysis of different methods for measuring baseline metabolic tumour volume

Luca Ceriani[1,2] · Lisa Milan[1] · Peter W. M. Johnson[3] · Maurizio Martelli[4] · Stefano Presilla[5] · Luca Giovanella[1] · Emanuele Zucca[2,6,7]

Abstract

Purpose This study assessed the performance of four different methods for the estimation of metabolic tumour volume (MTV) in primary mediastinal B cell lymphoma (PMBCL).

Method MTV was estimated using either a region growing automatic software program (RG) or a fixed threshold (FT) segmentation algorithm with the three most common cut-offs proposed in the literature (i.e., 25% and 41% of the SUVmax and SUV value ≥2.5). We compared these four methods using phantoms that simulated different set-ups of the main imaging characteristics of PMBCL (volume, shape, 18-FDG uptake and intra-lesion distribution) and assessed their performance in 103 PMBCL patients enrolled in the International Extranodal Lymphoma Study Group-26 (IELSG-26) study.

Results There was good correlation between MTV values estimated in vitro and in vivo using the different methods. The 25% FT cut-off (FT25%) provided the most accurate MTV evaluation in the phantoms. The cut-off at SUV 2.5 (FT2.5) resulted in MTV overestimation that particularly increased with high SUV values. The 41% cut-off (FT41%) showed MTV underestimation that was more evident when there were high levels of heterogeneity in tracer distribution. Shape of the lesion did not affect MTV computation. The RG algorithm provided a systematic slight MTV underestimation without significant changes due to lesion characteristics. We observed analogous trends for the MTV estimation in patients, with very different derived thresholds for the four methods. Optimal cut-offs for predicting progression-free survival (PFS) ranged from 213 to 831 ml. All methods predicted PFS with similar negative predictive values (94–95%) but different positive predictive values (23–45%).

Conclusions The different methods result in significantly different MTV cut-off values. All allow risk stratification in PMBCL, but FT25% showed the best capacity to predict disease progression in the patient cohort and provided the best accuracy in the phantom model.

Keywords Positron-emission tomography · Lymphoma · PET metrics · Metabolic tumour volume · Prognosis · Imaging

---

Luca Ceriani and Lisa Milan contributed equally to this work.

✉ Luca Ceriani
luca.ceriani@eoc.ch

[1] Nuclear Medicine and PET-CT Centre, Oncology Institute of Southern Switzerland, Via Ospedale 12, 6500 Bellinzona, Switzerland

[2] Institute of Oncology Research, Bellinzona, Switzerland

[3] Cancer Research UK Centre, Southampton General Hospital, Southampton, UK

[4] Department of Cellular Biotechnologies and Hematology, Sapienza University, Rome, Italy

[5] Medical Physics Unit, Ente Ospedaliero Cantonale, Bellinzona, Switzerland

[6] Division of Medical Oncology, Oncology Institute of Southern Switzerland, Bellinzona, Switzerland

[7] Department of Medical Oncology, Inselspital / Bern University Hospital, Bern, Switzerland

Introduction

Primary mediastinal (thymic) large B cell lymphoma (PMBCL) is a distinct entity arising from B cells within the thymus [1–5]. It usually presents with a single, large anterior mediastinal mass, often with local invasion and compressive syndromes [6]. PMBCL commonly affects young adults, and treatment with aggressive immunochemotherapy regimens, with or without radiotherapy, appears to yield good outcomes, with 5-year survival rates of over 90% in recent studies [7, 8]. However, salvage treatment for the few patients in whom initial therapy fails, generally has a poor outcome [9, 10]. Since early identification of the minority of high-risk patients would allow the development of risk-stratified approaches, there is an unmet need for reliable prognostic markers.

There is increasing evidence of the prognostic value of quantitative parameters obtained from 18FDG-PET/CT imaging in patients with aggressive non-Hodgkin lymphoma [11–13]. Standardized uptake value (SUV) has previously been the most widely studied parameter [14–16]. More recently, there has been growing recognition of volume-based metabolic assessment, including metabolic tumour volume (MTV) and total lesion glycolysis (TLG; defined as the product of MTV and SUVmean) as promising prognostic indices in solid tumours [17], multiple myeloma [18] and malignant lymphomas [19–28]. A number of different methods for the calculation of volume based PET parameters have been proposed. In lymphomas, the 3D isocontour method using a fixed threshold as cut off has been widely applied, but the definition of the best cut-off is still a matter of debate.

Meignan et al. tested a fixed threshold of 41% SUVmax, demonstrating good correlation between the segmented volumes and actual dimensions of the lesions in a phantom model, simulating the characteristics of the more common disseminated lymphomas [29]. This cut off, recommended by EANM guidelines to segment tumour volume [30], has been validated by the same group as prognostic in Hodgkin lymphoma and diffuse large B cell lymphoma [29, 31]. It has also been adopted for the study of other subtypes, including follicular lymphoma and peripheral T cell lymphoma [27].

More recently, Ilyas et al. compared different methods of MTV calculation in DLBCL (including the fixed threshold of 41% of SUVmax), showing that a SUV threshold of 2.5 achieved the best inter-observer agreement and was easiest to apply, although all methods predicted progression free (PFS) and overall survival (OS) with similar accuracy [32].

Although these studies included patients with bulky disease, ranging from 9 to 40% of cases, the results obtained in disseminated lymphomas cannot be directly applied in PMBCL where the tumor burden is characterized in >50% of patients by a single mass with diameter > 10 cm and in >80% by a lesion with diameter > 7 cm, respectively [8]. We have previously shown that a threshold at 25% performs better than the standard approaches for segmentation of the bulky mediastinal lesions in PMBCL and that MTV and TLG, estimated with this threshold, are powerful predictors of clinical outcome [33].

Therefore, in this study we compared MTV estimations obtained using three standard fixed threshold (FT) segmentation algorithms (differing only in the applied cut-off) and a contrast-based automatic region growing (RG) algorithm. We used an anthropomorphic phantom simulating the particular clinical presentation of PMBCL to investigate how large, very FDG-avid masses may affect the performance of these different methods. We then tested the same methods in a cohort of PMBCL patients to assess their performance in real mediastinal lesions and to compare their accuracy in predicting PFS and OS in this type of lymphoma.

Materials and methods

Phantom

To simulate the characteristics of the mediastinal lesions found in patients with PMBCL, we assembled a simplified anthropomorphic phantom reproducing the main features of the disease (Fig. 1). The morpho-metabolic characteristics of the mediastinal lesions and the other intra-thoracic structures in the phantoms were derived from those estimated in 103 PMBCL patients enrolled in the IELSG-26 study (NCT00944567) [8]. In this cohort, the MTV was calculated with a cut-off of 25% of the SUVmax [33] and the intra-tumour heterogeneity of tracer uptake was expressed as the area under the curve (AUC) of the cumulative SUV volume histograms (CSH) (AUC-CSH) [34].

Four series of phantoms were assembled with variable volume, shape (regular/irregular), FDG uptake, and FDG distribution patterns (uniform/heterogeneous) chosen to represent the range estimated in IELSG-26 patients. All the SUV values in the range were tested; but for volume and heterogeneity the choice of the values was partially conditioned by technical reasons since the fixed structure and dimensions of the inserts did not allow to include lesions with very low volume or very high AUC-CSH in the phantom model (Table 1).

All phantoms were scanned with a hybrid PET-CT scanner (MCT40, Siemens AG, Germany) and images were reconstructed using a standard clinical protocol with CT-based attenuation correction. The Online Resource summarises the main phantom characteristics and provides further methodological details (Table S1).

Volume segmentation

Volumes were measured on PET/CT acquisitions using a dedicated FDA approved software program (TrueD-Syngo,

Fig. 1 Characteristics of the assembled anthropomorphic phantom. **A** Scheme of the anthropomorphic phantom (regular shape and homogeneous uptake settings of the phantom lesion) with mediastinal lesion [1], cardiac insert [2] and mediastinal great vessels insert [3]. **B**: PET-CT image volumetric reconstruction of the phantom and fusion. **C** Phantom with lesion insert simulating an irregularly shaped mediastinal mass. **D** Insert assembled to simulate mediastinal mass with inhomogeneous intra-lesion tracer distribution
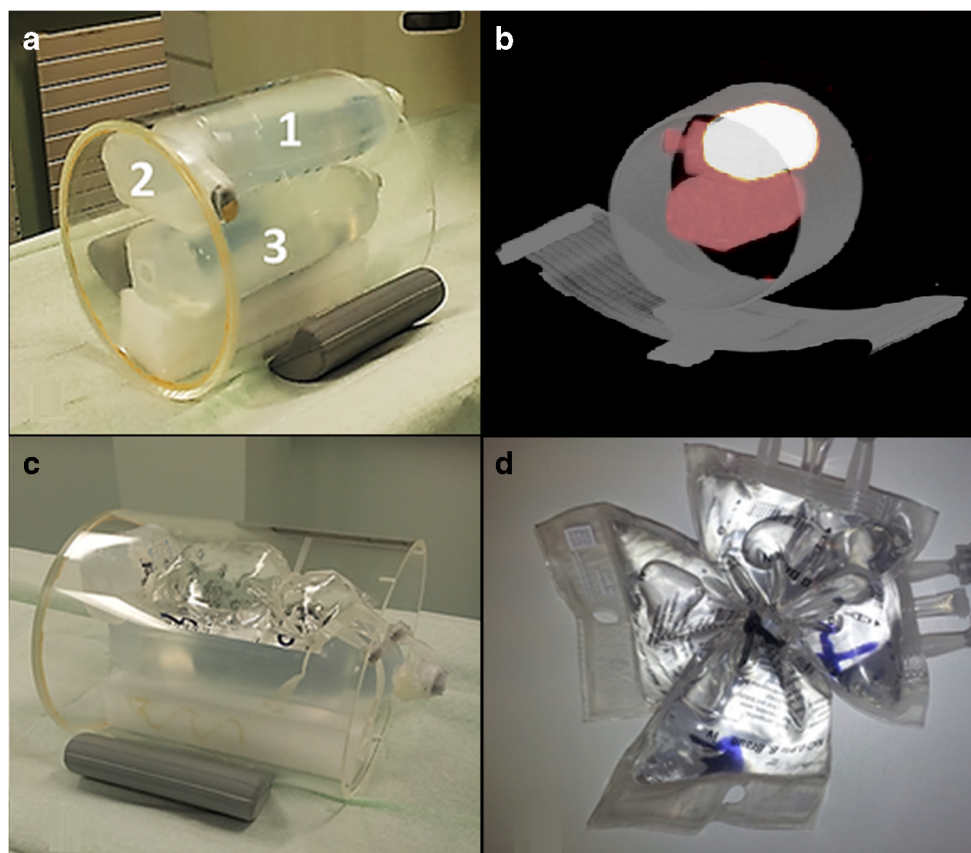
Siemens AG, Germany). A spherical box was positioned around the phantom insert mimicking the tumoral lesion to define a volume of interest (VOI). The MTV was then automatically calculated by a standardized routine program applying a fixed threshold (FT) algorithm. The tumour volume was delineated by all the voxels within the initial VOI with SUV values above or equal to the fixed threshold [35]. We tested the methods by applying three thresholds recently proposed in the literature for lymphomas: 25%, 41% of the SUVmax and an absolute SUV value of 2.5 (FT25%, FT41%, FT2.5) [29, 30, 33, 36]. In addition, the MTV calculation was also obtained by means of a region growing (RG) algorithm (3D Freeform Isocontour tool, TrueD Syngo, Siemens AG, Germany), with automatically segmented volume including all the voxels with common characteristics [37]. In PET images, after the selection of the tumor central voxels, all the neighboring voxels were automatically included in the segmented volume according to the voxel intensity. The algorithm was applied in contrast mode not requiring a predefined threshold [38].

Patients

Baseline PET scans of 103 patients with histologically proven PMBCL enrolled in the IELSG-26 study [8, 33] were analyzed with the same protocol and the methodology applied for the phantom imaging.

Statistical analysis

Quantitative variables were expressed as median and interquartile range (IQR). Medians were compared using the Wilcoxon signed-rank test. Absolute and relative differences between MTV measurement were expressed as mean and standard deviation (SD). Association between two variables was investigated through regression and correlation analysis (Pearson's coefficient). Agreement between methods was quantified with the intra-class correlation coefficient (ICC). Survival functions were defined according to the revised NCI criteria [12] and estimated by the Kaplan-Meier method; patient groups were compared by the log-rank test. Cox proportional hazard models were used for the estimation of hazard ratios (HRs). The ability of different methods to correctly identify patients at risk of shorter survival was compared using both the Uno's C concordance probability estimator (CPE) [39] and an in-sample fit approach to model selection (Akaike information criterion, AIC) [40]. A p value <0.05 was considered statistically significant. Negative predictive value (NPV) and positive predictive value (PPV) were calculated according to standard definitions [41]. Statistical analyses were conducted by using the STATA statistical software package, version 11 (StataCorp, College Station, TX) and the R statistical software environment, version 3.1.1 or the

Table 1 Characteristics of the mediastinal lesions and the other intra-thoracic structures in the PMBCL patients enrolled in the IELSG-26 study and the corresponding values used in the assembled phantoms

Characteristic	Patients			Phantoms
	Range	Median	IQR	Range
Mediastinal lesions				
SUVmax	8.3–36.6	18.8	15.5–22.9	10–36
MTV (ml)*	106–1271	406	267–640	193–1256
Heterogeneity (AUC-CSH)	0.32–0.65	0.49	0.43–0.54	0.43–0.65
Intra-thoracic vascular structures				
SUVmax MBP	1.18–3.02	1.87	1.58–2.16	1.7–1.9

Legend: AUC, area under the curve; CSH, cumulative SUV volume histogram; IQR, interquartile range; MBP, mediastinal blood pool; MTV, metabolic tumour volume; SUVmax, maximum standardized uptake value;*MTV was estimated using a fixed threshold at 25% of SUVmax

MedCalc Statistical Software, version 15.8 (MedCalc software bvba, Ostend; Belgium), as appropriate.

Results

Phantom models

Lesions of increasing volume (from 193 to 1256 ml) with fixed uptake and homogeneous tracer distribution (Fig. 2A)

Among the tested cut-offs, the FT25% provided the most accurate estimation of the actual volume (Online Resource, Table S1). The results obtained with this cut-off showed a consistent small overestimation of MTV, independent of the lesion volume (mean ± SD, 2.58% ± 0.9%). The RG method showed a systematic slight underestimation (mean ± SD, −3.83% ± 0.7%). The use of the higher 41% threshold resulted in underestimation of the lesion volumes (mean ± SD, −4.9% ± 1.8) with a progressive worsening of performance for the smallest lesions (≤401 ml). On the other hand, the FT2.5 showed a clear overestimation (13.85% ± 4.0), which was reduced for the smallest lesions.

Lesions of increasing heterogeneity of tracer distribution (AUC-CSH from 0.66 to 0.43) with fixed volume and fixed uptake (Fig. 2B)

For FT41%, a rise in heterogeneity caused a progressive volume underestimation in comparison with the results obtained with the homogeneous insert, exceeding 15% for the greatest heterogeneity. The FT25% provided the most accurate MTV calculation (error of −0.49% for the lesion with highest heterogeneity), although there was a slight inverse correlation between volume estimated and degree of heterogeneity. FT2.5 and RG MTV estimation were not affected by the degree of heterogeneity.

Lesions of increasing uptake (SUV value from 10 to 36) with fixed volume and homogeneous tracer distribution (Fig. 2C)

The results demonstrated that the uptake level of the lesion does not significantly affect the accuracy of methods using percentage threshold (FT25% and FT41%) and the RG program. Conversely, FT2.5, applying a fixed absolute threshold, showed a progressive over-estimation of MTV correlated to the SUV values.

Lesions of fixed volume with irregular shape and homogeneous uptake

The four methods estimated the MTV of an irregularly shaped lesion with accuracy comparable to that demonstrated for the lesions of regular shape and similar volume (error < 1%).

The Online Resource (Tables S2–S4) reports the detailed results of MTV estimation by the different segmentation methods (with their differences and relative errors with respect to the actual phantom volume).

Patient population

The median values of MTV estimated by the different approaches (FT25%, FT 41%, FT2.5, RG) are summarized in Table 2 which also reports the absolute and percentage thresholds of the SUV values applied by different methods. The RG algorithm segmented MTV applying thresholds ranging between 10% and 40% of SUVmax, with mean value slightly higher than 25% (mean = 27.5%, median = 28%, IQR 23–32%). The cut-off value of SUV 2.5 corresponded in the overall population to a median SUV max percentage of 13% (IQR 10.4–15.4%).

We observed statistically significant differences between patient MTVs estimated by the four methods (Wilcoxon signed-rank test $p < 0.0001$ for the four methods considered together and $p < 0.001$ for each pair of methods compared). These relative differences hinged on the different thresholds used (Table 2), and we found a highly significant ($p < 0.0001$)

Fig. 2 Relative error of MTV estimation by the different methods. The relative error of MTV estimation in the anthropomorphic phantom models using different methods with changing of volume (**A**), of uptake heterogeneity (**B**) and of uptake intensity of the phantom lesion (**C**)
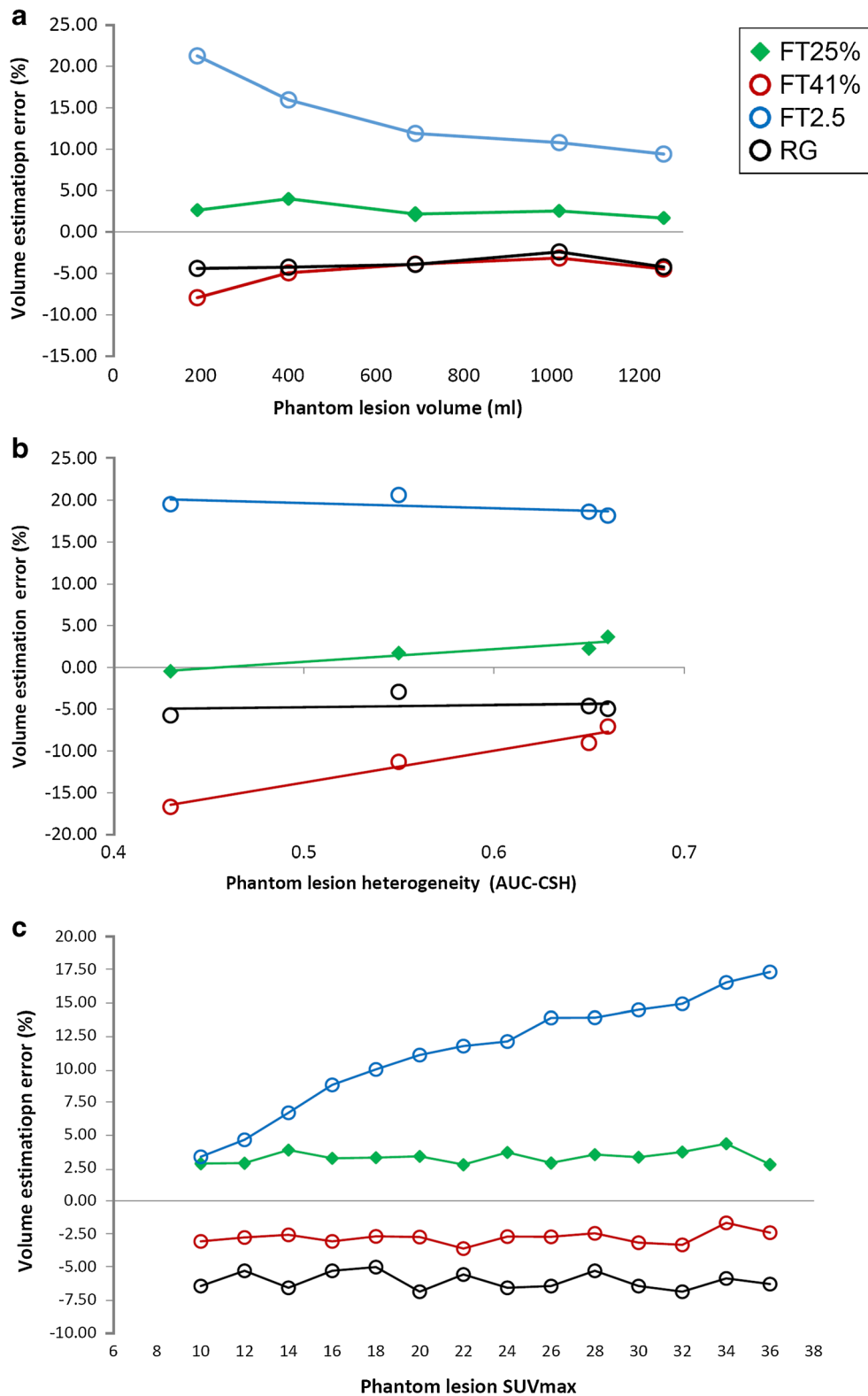
inverse correlation between median MTV and threshold SUV values (Online Resource, Fig. S1). In nearly all patients, the FT2.5 provided the highest MTV value, while FT41% contoured the smallest MTV (Online Resource, Fig. S2).

Nevertheless, there was an extremely good correlation between all methods for the MTV estimation: the Pearson's coefficient ranges between 0.8 and 0.9, with p values consistently <0.0001 (Online Resource, Table S5).

Table 2 Comparison of different segmentation methods with respect to applied thresholds and estimated metabolic tumour volumes, and analysis of relative differences with intra-class correlation

Parameter	Segmentation method			
	FT25%	FT2.5	RG	FT41%
Threshold applied				
Absolute SUV value	Median, 4.9	2.5	Median, 5.3	Median, 8
IQR	4.1–6.0	(−)	4.0-6.6	6.6–9.9
Percent of SUVmax	25%	Median, 13%	Median, 28%	41%
IQR	(−)	10.4–15.4%	23.0–32.0%	(−)
Estimate of MTV (ml)				
median	406	613	380	220
IQR	267–641	403–868	224–577	123–327
Min-max values	86–1271	129–1918	44–1247	9–814
Absolute difference (ml) vs. FT25%, mean ± SD	0	198 ± 175	−64 ± 122	−229 ± 174
Relative difference (%) vs. FT25%, mean ± SD	0	49 ± 39	−12 ± 19	−48 ± 19
ICC vs. FT25%	1	0.85	0.89	0.71

Legend: SUV, standardized uptake value; IQR, interquartile range; max, maximum; min, minimum; SD, standard deviation; SUVmax, maximum standardized uptake value; MTV, metabolic tumour volume; ICC, intra-class correlation; FT, fixed threshold; RG, region growing

The difference between FT25% and the other methods appears independent from the MTV value for FT41% and RG, while for FT2.5, it decreases progressively with increasing size of the lesion as demonstrated in the phantom model (Fig. 3A).

FT41% showed higher MTV underestimation in patients with greater heterogeneity, while the relative MTV overestimation by FT2.5 was higher in cases with higher SUVmax, again, in keeping with the trend observed in the phantom models (Fig. 3B and C).

Notably, the relative percentage differences of the MTV estimated by the different methods was 3–4 fold higher in patients than in phantoms (Fig. 3A).

The correlation between MTV measurements obtained by different methods (ICC), varied from 0.89 (between FT25% and RG) to 0.71 (between FT25% and FT41%) confirming the presence of inter-method differences (Table 2).

Prognostic value of MTV

The distribution and area under the ROC curves for all four methods were similar for prediction of either overall survival or progression-free survival, with the exception of a significantly different AUC of the ROC curves between FT25% and FT41% methods (representing the largest and smallest value, respectively) for the analysis of progression free survival. Although all the methods allowed the identification of patients with significantly higher risk of progression, they provided different MTV cut-points for the separation of low- and high-risk groups (Table 3). The optimal MTV cut-point for the FT2.5 method was approximately three times greater than the one derived from the FT41% method. The four methods

demonstrated similar negative NPV (from 94% to 95%) but different PPV (from 20% to 45%). Kaplan–Meier estimates (Fig. 4) showed that the patients with low MTV have a significantly longer PFS and OS and compared to the patients with high MTV, regardless of the method used. A concordance probability estimation (Uno's C method) and the Akaike global model fit criterion showed that FT25% achieved both the best fit (lower AIC) and the best discrimination (higher CPE) in our cohort to identify patients with shorter PFS and OFS (Online Resource, Table S6).

Additionally, elevated TLG, which is the product of MTV and SUVmean, was also significantly associated with a shorter PFS and OS in PMBCL, irrespective of the segmentation method adopted. Analogous to MTV, the best discrimination of patients at different risk of treatment failure was also provided using the FT25% method for the estimation of TLG (Online resource, Tables S7 and S8).

Discussion

There is increasing evidence, mainly described in large B cell lymphoma, but also in several other histologic lymphoma subtypes, of the prognostic value of quantitative parameters (SUVmax, MTV and TLG) obtained from initial staging with 18FDG-PET/CT [11, 12, 26, 33, 42]. However, their clinical use is still constrained by technical limitations, which make it difficult to obtain comparable results across various PET/CT systems and institutions. The cut-point that defines the patient with high MTV and poor risk disease is particularly inconsistent and depends on the method used for its estimation; its variability in different reports is also related to the small
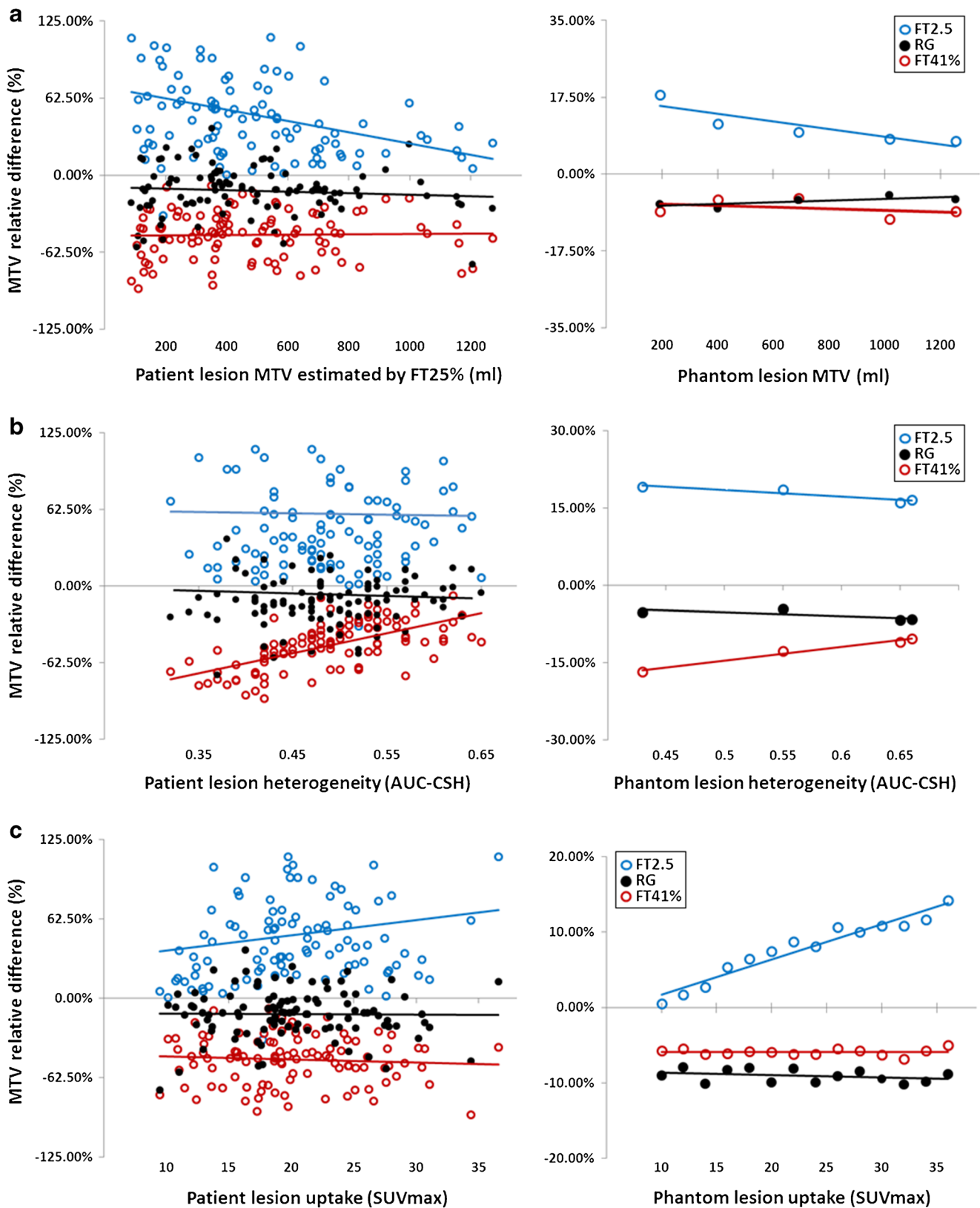
Fig. 3 Relative difference in MTV values estimated by different methods in the patients population *(left)* and in anthropomorphic phantom models *(right)*. The relative difference of MTV estimation by FT2.5, FT41% and RG methods with respect to FT25 is plotted as a function of lesion volume (**A**), heterogeneity (**B**) and uptake intensity of the phantom lesion (**C**)

Table 3 Description of the ROC analysis used to identify optimal cut-off points and outcome results according to the different segmentation methods

Survival	Method	ROC curve analysis					Kaplan-Meier curve analysis				
		MTV cut-off	Sens	Spec	AUC	P value	5-year survival probability		P value	NPV%	PPV%
							High MTV patients	Low MTV patients			
Progression Free Survival	FT25%	>703	0.69	0.88	0.814	<0.0001	0.95	0.55	<0.0001	95	45
	FT41%	>213	0.92	0.54	0.724	0.002	0.94	0.80	0.0253	94	20
	FT2.5	>831	0.77	0.79	0.786	<0.0001	0.95	0.64	<0.0001	95	36
	RG	>391	0.92	0.61	0.781	<0.0001	0.95	0.77	0.0081	95	23
Overall Survival	FT25%	>490	1.00	0.61	0.812	<0.0001	1.00	0.86	0.0035	100	14
	FT41%	>213	1.00	0.52	0.701	0.017	1.00	0.89	0.0148	100	11
	FT2.5	>657	1.00	0.58	0.781	0.0001	1.00	0.87	0.0060	100	13
	RG	>391	1.00	0.58	0.758	0.003	1.00	0.87	0.0060	100	13

Legend: ROC, receiver operating characteristic; MTV, metabolic tumour volume; FT, fixed threshold; RG, region growing; sens, sensitivity; spec, specificity; AUC, area under the curve; NPV, negative predictive value; PPV, positive predictive value

sample size of most individual studies. The present study addressed the need to optimize the MTV measure in patients with PMBCL, where baseline quantitative PET parameters have been shown to be powerful prognostic indicators [33] and might have a role in treatment decisions [34].

We developed a phantom model that attempted to reproduce the clinical and metabolic features observed in the largest prospective study on PMBCL patients [8]. Our phantom is, to the best of our knowledge, the first lymphoma model that, in addition to volume, shape and uptake, also takes into account the metabolic tumour heterogeneity, which we recently showed to be helpful for the early identification of poor-risk PMBCL [34]. We tested performance of the different methods (FT25%, FT41%, FT2.5, RG) proposed in the literature for the metabolic volume segmentation in various phantom models.

Our results showed that FT25% provides the best concordance between measured and actual volumes, with a performance only slightly affected by the lesion characteristics and maintained when metabolic heterogeneity is included in the models to make phantoms more adherent to the biological features of the disease. These experimental data are consistent with the clinical observations we made in prior studies on the use of PET in PMBCL patients [8, 33], in which the 25% threshold, selected by qualitative visual analysis of the best fit between metabolic edges and anatomic boundaries, was superior to the 41% currently recommended for volume segmentation in lymphoma patients [30]. In fact, the particular clinical features of PMBCL, usually characterized by bulky mediastinal mass, limit the potential advantages of the 41% cut-off, which was developed in diffuse large B cell lymphoma (DLBCL), which often presents with multiple lesions at different sites, of variable size, background and metabolic uptake.

In phantom models, FT2.5 led to a systematic volume overestimation. Conversely, FT41% underestimated the volume of

large FDG-avid masses, particularly in highly heterogeneous lesions. The RG algorithm, which applies a variable threshold, also showed a generalized volume underestimation. These phantom model results appear in keeping with those obtained in patients with PMBCL. The relative percentage differences of the MTV estimated in patients by the different methods have been consistently higher than those obtained in the PET images of the phantoms. These findings may be explained by the higher variability of volume, tracer uptake, and heterogeneity in tumor lesions in the patient population compared with the phantom model. Although it is not possible to specify in individual patients the weight of each variable, the trends that we observed in patients in the relationship between segmentation methods and lesion characteristics appeared similar to those demonstrated in phantom models.

We demonstrated that in PMBCL, defined by a single large lesion, the estimated MTV values are strongly method-related and the four approaches provided significantly different MTV values. Also the optimal cut-offs for MTV to predict PFS and OS were very different and method-related, as recently shown in DLBCLs [32]. While these inter-method variations are expected, the finding, by ROC curve analysis, of different intra-method cut-offs for PFS and OS is difficult to explain on the basis of biology. This limitation may be overcome by the analysis of larger cohorts, such as the one of the ongoing IELSG37 study (NCT NCT01599559), which aims to enroll more than 500 patients.

All the methodologies demonstrated predictive power in risk-stratifying mediastinal masses, confirming the intrinsic prognostic role of baseline MTV in PMBCL. Although all

Fig. 4 Risk group discrimination in PMBCL patients according to MTV estimated by different methods. Kaplan-Meier curves of PFS (panel A) and OS (panel B) according to MTV at diagnosis estimated using different segmentation algorithms
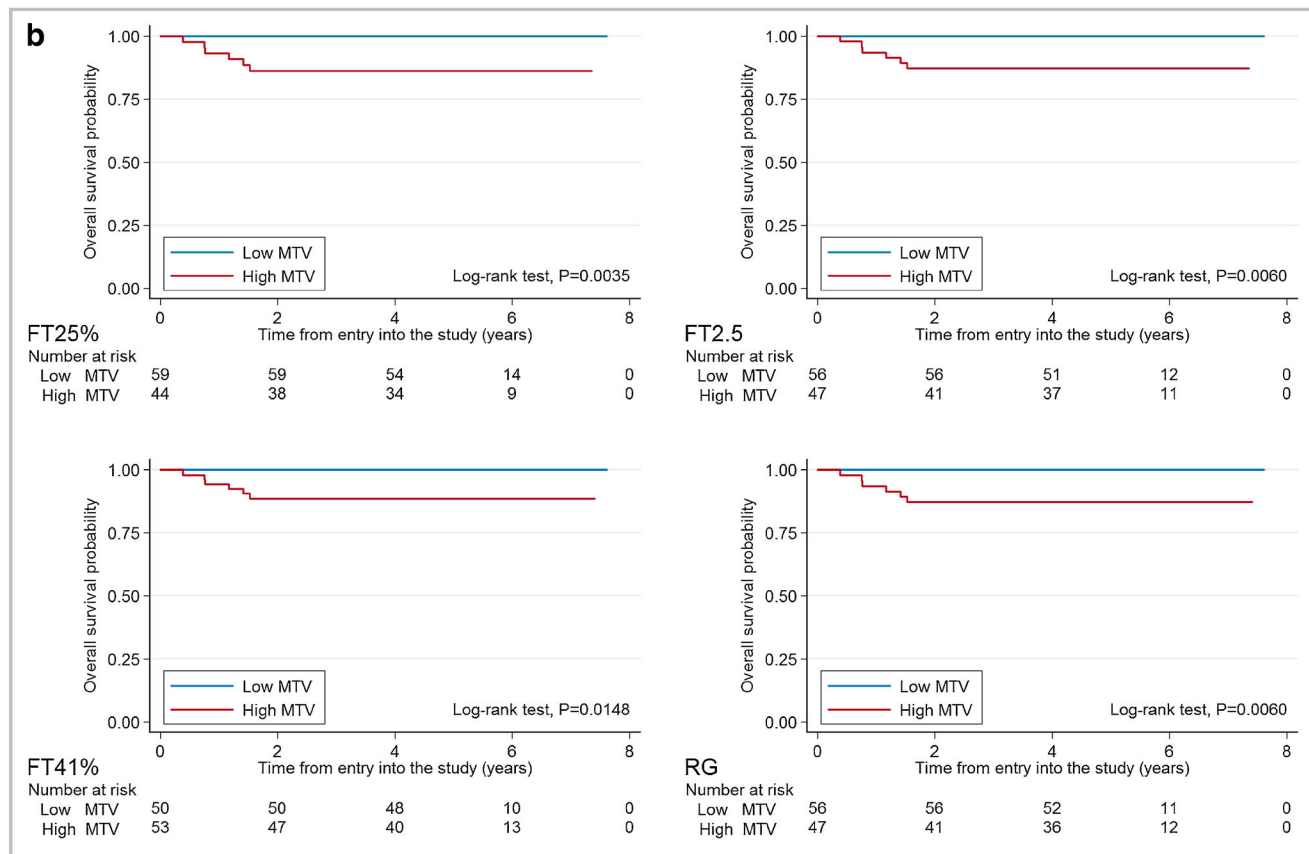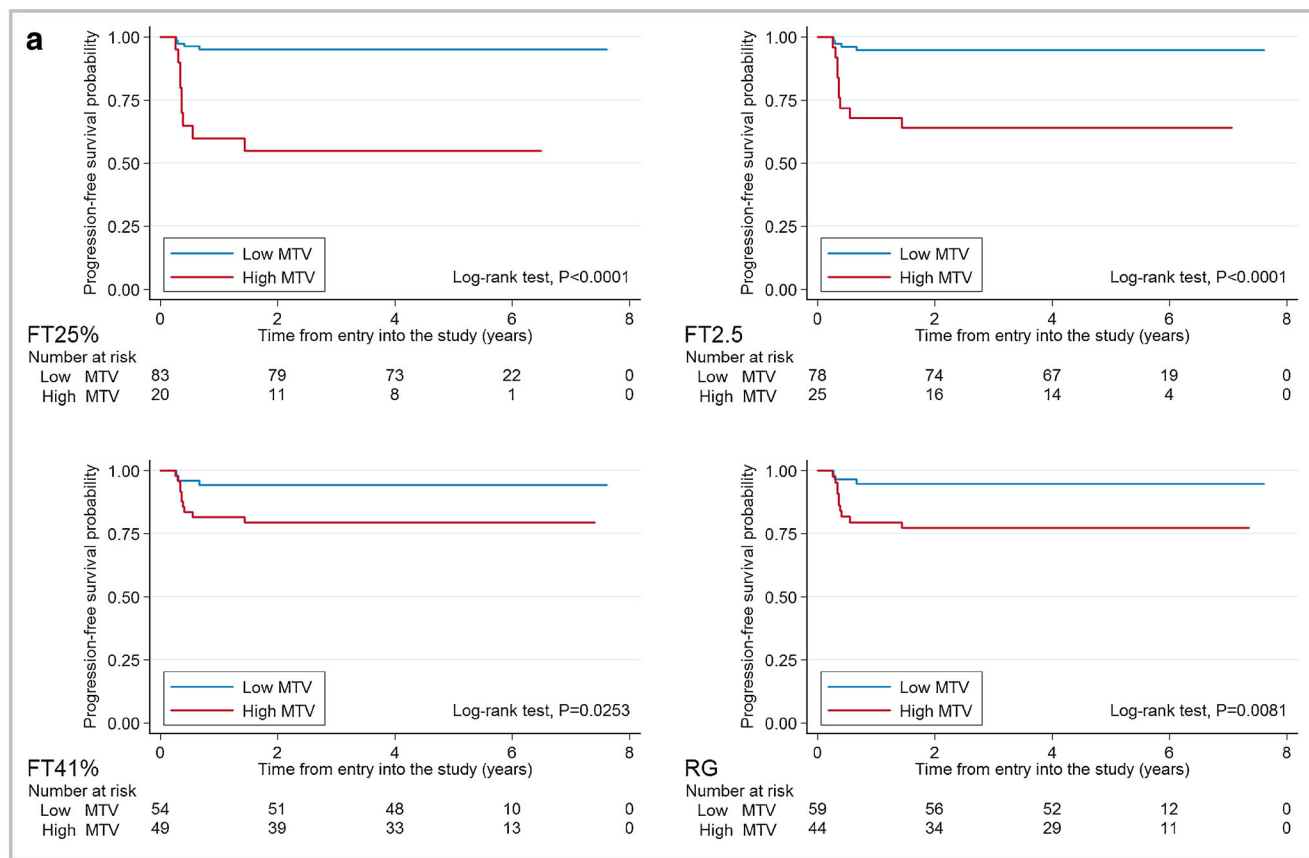
four methods could predict PFS with good accuracy and similar NPV (95–98%) the FT25% showed the highest PPV (45%) and provided the best model to identify patients with poorer outcomes in this population.

Our results confirm the relationship between the most accurate estimation of the actual dimensions of the lesions and the best risk classification. We found individual patients in the study with very high FDG uptake and others with very heterogeneous masses, where FT41% and FT2.5 appeared to either under- or over-estimate MTV in comparison with FT25%.

Recently, Ilyas et al. proposed FT2.5 as the preferred method to segment PET volumes in patients with DLBCL because the use of an absolute cut-off (i.e. SUV =2.5) would be less time consuming, more reproducible and easier to apply than the other methods based on a percentage of the maximum SUV (i.e. 41%).

These considerations are of less importance in PMBCL, which is characterized by one dominant lesion. In this specific situation, the more common standard programs, after drawing the initial volume of interest around the pathological 18FDG uptake, automatically define the edges of the lesion in a single step, applying the chosen cut-off, either an absolute SUV value or a percentage of the maximum SUV. Our experience using FT25% in this setting showed very high inter-observer reproducibility (Pearson correlation coefficient > 0.99 and $P < 0.0001$) for the calculation of MTV [33].

Our prior studies have shown that, among functional PET parameters, elevated TLG is the best individual predictor of poorer PFS [33, 34]. In the present study, analogous to what was described for MTV, TLG values calculated with the FT25% method provided the best PPV, although all segmentation methods were able to correctly identify patients with favorable outcome.

In conclusion, this study has established that a cut-off of 25% SUVmax of the lesion is the optimal method for measuring baseline MTV in PMBCL and provides the best discrimination of patient risk.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in this study were in accordance with the ethical standards of the 1964 Helsinki declaration and its later amendments.

The IELSG-26 study was authorized at the Oncology Institute of Southern Switzerland by the local Cantonal Ethics Committee (CE-TI-1772) and approved by the Institutional Review Board/Ethics Committee of each participating centre.

Informed consent All patients enrolled provided written informed consent for the use of their anonymized imaging and clinical data.

References

1. Boleti E, Johnson PW. Primary mediastinal B-cell lymphoma. Hematol Oncol. 2007;25:157–63. https://doi.org/10.1002/hon.818.
2. Gaulard P, Harris NL, Pileri SA, Kutock JL, Stein H, Kovrigina AM, et al. Primary mediastinal (thymic) large B-cell lymphoma. In: Swerdlow S, Campo E, Harris NL, Jaffe ES, Pileri SA, Stein H, editors. WHO classification of tumours of haematopoietic and lymphoid tissues. Lyon: IARC; 2008. p. 250–3.
3. Steidl C, Gascoyne RD. The molecular pathogenesis of primary mediastinal large B-cell lymphoma. Blood. 2011;118:2659–69. https://doi.org/10.1182/blood-2011-05-326538.
4. Savage KJ, Monti S, Kutok JL, Cattoretti G, Neuberg D, De Leval L, et al. The molecular signature of mediastinal large B-cell lymphoma differs from that of other diffuse large B-cell lymphomas and shares features with classical Hodgkin's lymphoma. Blood. 2003;102:3871–9. https://doi.org/10.1182/blood-2003-06-1841.
5. Rosenwald A, Wright G, Leroy K, Yu X, Gaulard P, Gascoyne RD, et al. Molecular diagnosis of primary mediastinal B cell lymphoma identifies a clinically favorable subgroup of diffuse large B cell lymphoma related to Hodgkin lymphoma. J Exp Med. 2003;198:851–62. https://doi.org/10.1084/jem.20031074.
6. Johnson PW, Davies AJ. Primary mediastinal B-cell lymphoma. Am Soc Hematol Educ Program. 2008;2008:349–58. https://doi.org/10.1182/asheducation-2008.1.349.
7. Dunleavy K, Pittaluga S, Maeda LS, Advani R, Chen CC, Hessler J, et al. Dose-adjusted EPOCH-rituximab therapy in primary mediastinal B-cell lymphoma. N Engl J Med. 2013;368:1408–16. https://doi.org/10.1056/NEJMoa1214561.
8. Martelli M, Ceriani L, Zucca E, Zinzani PL, Ferreri AJ, Vitolo U, et al. [18F]fluorodeoxyglucose positron emission tomography predicts survival after chemoimmunotherapy for primary mediastinal large B-cell lymphoma: results of the international extranodal lymphoma study group IELSG-26 study. J Clin Oncol. 2014;32:1769–75. https://doi.org/10.1200/jco.2013.51.7524.
9. Kuruvilla J, Pintilie M, Tsang R, Nagy T, Keating A, Crump M. Salvage chemotherapy and autologous stem cell transplantation are inferior for relapsed or refractory primary mediastinal large B-cell lymphoma compared with diffuse large B-cell lymphoma. Leuk Lymphoma. 2008;49:1329–36. https://doi.org/10.1080/10428190802108870.
10. Todeschini G, Secchi S, Morra E, Vitolo U, Orlandi E, Pasini F, et al. Primary mediastinal large B-cell lymphoma (PMLBCL): long-term results from a retrospective multicentre Italian experience in 138 patients treated with CHOP or MACOP-B/VACOP-B. Br J Cancer. 2004;90:372–6. https://doi.org/10.1038/sj.bjc.6601460.
11. Cheson BD. Role of functional imaging in the management of lymphoma. J Clin Oncol. 2011;29:1844–54. https://doi.org/10.1200/JCO.2010.32.5225.
12. Cheson BD, Pfistner B, Juweid ME, Gascoyne RD, Specht L, Horning SJ, et al. Revised response criteria for malignant lymphoma. J Clin Oncol. 2007;25:579–86.
13. Biggi A, Gallamini A, Chauvie S, Hutchings M, Kostakoglu L, Gregianin M, et al. International validation study for interim PET in ABVD-treated, advanced-stage hodgkin lymphoma:

interpretation criteria and concordance rate among reviewers. J Nucl Med. 2013;54:683–90. https://doi.org/10.2967/jnumed.112.110890.

14. Itti E, Meignan M, Berriolo-Riedinger A, Biggi A, Cashen AF, Vera P, et al. An international confirmatory study of the prognostic value of early PET/CT in diffuse large B-cell lymphoma: comparison between Deauville criteria and DeltaSUVmax. Eur J Nucl Med Mol Imaging. 2013;40:1312–20. https://doi.org/10.1007/s00259-013-2435-6.

15. Itti E, Lin C, Dupuis J, Paone G, Capacchione D, Rahmouni A, et al. Prognostic value of interim 18F-FDG PET in patients with diffuse large B-cell lymphoma: SUV-based assessment at 4 cycles of chemotherapy. J Nucl Med. 2009;50:527–33. https://doi.org/10.2967/jnumed.108.057703.

16. Salles G, Seymour JF, Offner F, Lopez-Guillermo A, Belada D, Xerri L, et al. Rituximab maintenance for 2 years in patients with high tumour burden follicular lymphoma responding to rituximab plus chemotherapy (PRIMA): a phase 3, randomised controlled trial. Lancet. 2011;377:42–51. https://doi.org/10.1016/s0140-6736(10)62175-7.

17. Van de Wiele C, Kruse V, Smeets P, Sathekge M, Maes A. Predictive and prognostic value of metabolic tumour volume and total lesion glycolysis in solid tumours. Eur J Nucl Med Mol Imaging. 2013;40:290–301. https://doi.org/10.1007/s00259-012-2280-z.

18. Fonti R, Larobina M, Del Vecchio S, De Luca S, Fabbricini R, Catalano L, et al. Metabolic tumor volume assessed by 18F-FDG PET/CT for the prediction of outcome in patients with multiple myeloma. J Nucl Med. 2012;53:1829–35. https://doi.org/10.2967/jnumed.112.106500.

19. Kim CY, Hong CM, Kim DH, Son SH, Jeong SY, Lee SW, et al. Prognostic value of whole-body metabolic tumour volume and total lesion glycolysis measured on (1)(8)F-FDG PET/CT in patients with extranodal NK/T-cell lymphoma. Eur J Nucl Med Mol Imaging. 2013;40:1321–9. https://doi.org/10.1007/s00259-013-2443-6.

20. Song MK, Chung JS, Shin HJ, Lee SM, Lee SE, Lee HS, et al. Clinical significance of metabolic tumor volume by PET/CT in stages II and III of diffuse large B cell lymphoma without extranodal site involvement. Ann Hematol. 2012;91:697–703. https://doi.org/10.1007/s00277-011-1357-2.

21. Yang DH, Ahn JS, Byun BH, Min JJ, Kweon SS, Chae YS, et al. Interim PET/CT-based prognostic model for the treatment of diffuse large B cell lymphoma in the post-rituximab era. Ann Hematol. 2013;92:471–9. https://doi.org/10.1007/s00277-012-1640-x.

22. Manohar K, Mittal BR, Bhattacharya A, Malhotra P, Varma S. Prognostic value of quantitative parameters derived on initial staging 18F-fluorodeoxyglucose positron emission tomography/computed tomography in patients with high-grade non-Hodgkin's lymphoma. Nucl Med Commun. 2012;33:974–81. https://doi.org/10.1097/MNM.0b013e32835673ec.

23. Gallicchio R, Mansueto G, Simeon V, Nardelli A, Guariglia R, Capacchione D, et al. F-18 FDG PET/CT quantisation parameters as predictors of outcome in patients with diffuse large B-cell lymphoma. Eur J Haematol. 2014. https://doi.org/10.1111/ejh.12268.

24. Esfahani SA, Heidari P, Halpern EF, Hochberg EP, Palmer EL, Mahmood U. Baseline total lesion glycolysis measured with (18)F-FDG PET/CT as a predictor of progression-free survival in diffuse large B-cell lymphoma: a pilot study. Am J Nucl Med Mol Imaging. 2013;3:272–81.

25. Cazaentre T, Morschhauser F, Vermandel M, Betrouni N, Prangere T, Steinling M, et al. Pre-therapy 18F-FDG PET quantitative parameters help in predicting the response to radioimmunotherapy in non-Hodgkin lymphoma. Eur J Nucl Med Mol Imaging. 2010;37:494–504. https://doi.org/10.1007/s00259-009-1275-x.

26. Meignan M, Cottereau AS, Versari A, Chartier L, Dupuis J, Boussetta S, et al. Baseline metabolic tumor volume predicts outcome in high-tumor-burden follicular lymphoma: a pooled analysis of three multicenter studies. J Clin Oncol. 2016;34:3618–26. https://doi.org/10.1200/jco.2016.66.9440.

27. Cottereau AS, Becker S, Broussais F, Casasnovas O, Kanoun S, Roques M, et al. Prognostic value of baseline total metabolic tumor volume (TMTV0) measured on FDG-PET/CT in patients with peripheral T-cell lymphoma (PTCL). Ann Oncol. 2016;27:719–24. https://doi.org/10.1093/annonc/mdw011.

28. Kanoun S, Rossi C, Berriolo-Riedinger A, Dygai-Cochet I, Cochet A, Humbert O, et al. Baseline metabolic tumour volume is an independent prognostic factor in Hodgkin lymphoma. Eur J Nucl Med Mol Imaging. 2014;41:1735–43. https://doi.org/10.1007/s00259-014-2783-x.

29. Meignan M, Sasanelli M, Casasnovas RO, Luminari S, Fioroni F, Coriani C, et al. Metabolic tumour volumes measured at staging in lymphoma: methodological evaluation on phantom experiments and patients. Eur J Nucl Med Mol Imaging. 2014;41:1113–22. https://doi.org/10.1007/s00259-014-2705-y.

30. Boellaard R, Delgado-Bolton R, Oyen WJ, Giammarile F, Tatsch K, Eschner W, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. Eur J Nucl Med Mol Imaging. 2015;42:328–54. https://doi.org/10.1007/s00259-014-2961-x.

31. Sasanelli M, Meignan M, Haioun C, Berriolo-Riedinger A, Casasnovas RO, Biggi A, et al. Pretherapy metabolic tumour volume is an independent predictor of outcome in patients with diffuse large B-cell lymphoma. Eur J Nucl Med Mol Imaging. 2014;41:2017–22. https://doi.org/10.1007/s00259-014-2822-7.

32. Ilyas H, Mikhaeel NG, Dunn JT, Rahman F, Moller H, Smith D, 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;45:1142–54. https://doi.org/10.1007/s00259-018-3953-z.

33. Ceriani L, Martelli M, Zinzani PL, Ferreri AJ, Botto B, Stelitano C, et al. Utility of baseline 18FDG-PET/CT functional parameters in defining prognosis of primary mediastinal (thymic) large B-cell lymphoma. Blood. 2015;126:950–6. https://doi.org/10.1182/blood-2014-12-616474.

34. Ceriani L, Milan L, Martelli M, Ferreri AJM, Cascione L, Zinzani PL, et al. Metabolic heterogeneity on baseline 18FDG-PET/CT scan is a predictor of outcome in primary mediastinal B-cell lymphoma. Blood. 2018;132:179–86. https://doi.org/10.1182/blood-2018-01-826958.

35. Boellaard R, Krak NC, Hoekstra OS, Lammertsma AA. Effects of noise, image resolution, and ROI definition on the accuracy of standard uptake values: a simulation study. J Nucl Med. 2004;45:1519–27.

36. Firouzian A, Kelly MD, Declerck JM. Insight on automated lesion delineation methods for PET data. EJNMMI Res. 2014;4:69. https://doi.org/10.1186/s13550-014-0069-8.

37. Pratt WK. Image segmentation. Digital image processing: PIKS scientific inside, fourth edition. Wiley; 2007.

38. Vorwerk H, Beckmann G, Bremer M, Degen M, Dietl B, Fietkau R, et al. The delineation of target volumes for radiotherapy of lung cancer patients. Radiother Oncol. 2009;91:455–60. https://doi.org/10.1016/j.radonc.2009.03.014.

39. Gönen M, Heller G. Concordance probability and discriminatory power in proportional hazards regression. Biometrika. 2005;92:965–70. https://doi.org/10.1093/biomet/92.4.965.

40. Posada D, Buckley TR. Model selection and model averaging in Phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Syst Biol. 2004;53:793–808. https://doi.org/10.1080/10635150490522304.

41. Altman DG, Bland JM. Diagnostic tests 2: predictive values. BMJ (Clin Res ed). 1994;309:102.

42. Meignan M, Gallamini A, Haioun C. Report on the first international workshop on interim-PET-scan in lymphoma. Leuk Lymphoma. 2009;50:1257–60. https://doi.org/10.1080/10428190903040048.