

SAKK38/07 study: integration of baseline metabolic heterogeneity and metabolic tumor volume in DLBCL prognostic model

Luca Ceriani,^{1,2} Giuseppe Gritti,³ Luciano Cascione,^{2,4} Maria Cristina Piroso,⁵ Angela Polino,⁵ Teresa Ruberto,¹ Anastasios Stathis,⁵ Andrea Bruno,⁶ Alden A. Moccia,⁵ Luca Giovannella,^{1,7} Stefanie Hayoz,⁸ Sämi Schär,⁸ Stefan Dirnhofer,⁹ Alessandro Rambaldi,^{3,10} Giovanni Martinelli,¹¹ Christoph Mamot,¹² and Emanuele Zucca^{2,5,13}

¹Nuclear Medicine and PET/CT Centre, Imaging Institute of Southern Switzerland, Bellinzona, Switzerland; ²Institute of Oncology Research, Faculty of Biomedical Sciences, Università della Svizzera Italiana, Bellinzona, Switzerland; ³Hematology Unit, Azienda Ospedaliera Papa Giovanni XXIII, Bergamo, Italy; ⁴SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland; ⁵Medical Oncology Clinic, Oncology Institute of Southern Switzerland, Bellinzona, Switzerland; ⁶Department of Nuclear Medicine, Azienda Ospedaliera Papa Giovanni XXIII, Bergamo, Italy; ⁷Division of Nuclear Medicine, University Hospital and University of Zurich, Zurich, Switzerland; ⁸Swiss Group for Clinical Cancer Research (SAKK) Coordinating Center, Bern, Switzerland; ⁹Institute of Medical Genetics and Pathology, University Hospital Basel, University of Basel, Switzerland; ¹⁰Department of Oncology and Hemato-Oncology, University of Milan, Milan, Italy; ¹¹European Institute of Oncology, Milan, Italy; ¹²Cantonal Hospital Aarau, Aarau, Switzerland; and ¹³Department of Medical Oncology, Inselspital, University Hospital and University of Bern, Bern, Switzerland

Key Points

- Baseline functional PET parameters can predict outcomes of DLBCL patients treated with R-CHOP.
- Prognostic models based on MTV and MH may allow early identification of poor-risk patients.

Several functional parameters from baseline (18)F-fluorodeoxyglucose positron emission tomography (PET)/computed tomography have been proposed as promising biomarkers of treatment efficacy in diffuse large B-cell lymphoma (DLBCL). We tested their ability to predict outcome in 2 cohorts of DLBCL patients receiving conventional immunochemotherapy (rituximab, cyclophosphamide, doxorubicin hydrochloride, vincristine sulfate, and prednisone [R-CHOP] regimen), either every 14 (R-CHOP14) or 21 days (R-CHOP21). Baseline PET analysis was performed in 141 patients with DLBCL treated with R-CHOP14 in the prospective SAKK38/07 study (NCT00544219) of the Swiss Group for Clinical Cancer Research (testing set). Reproducibility was examined in a validation set of 113 patients treated with R-CHOP21. In the SAKK38/07 cohort, progression-free survival (PFS) at 5 years was 83% for patients with low metabolic tumor volume (MTV) and 59% for those with high MTV (hazard ratio [HR], 3.4; 95% confidence interval [CI], 1.6-7.0; $P = .0005$), whereas overall survival (OS) was 91% and 64%, respectively (HR, 4.4; 95% CI, 1.9-10; $P = .0001$). MTV was the most powerful predictor of outcome also in the validation set. Elevated metabolic heterogeneity (MH) significantly predicted poorer outcomes in the subgroups of patients with elevated MTV. A model integrating MTV and MH identified high-risk patients with shorter PFS (testing set: HR, 5.6; 95% CI, 1.8-17; $P < .0001$; validation set: HR, 5.6; 95% CI, 1.7-18; $P = .0002$) and shorter OS (testing set: HR, 9.5; 95% CI, 1.7-52; $P < .0001$; validation set: HR, 7.6; 95% CI, 2.0-28; $P = .0003$). This finding was confirmed by an unsupervised regression tree analysis indicating that prognostic models based on MTV and MH may allow early identification of refractory patients who might benefit from treatment intensification. This trial was registered at www.clinicaltrials.gov as #NCT00544219.

Introduction

Diffuse large B-cell lymphoma (DLBCL) not otherwise specified is the most common histologic subtype of lymphoma, accounting for 30% to 40% of the non-Hodgkin lymphomas in Western countries.¹ It is

a heterogeneous disease with distinct molecular features and clinical behavior.^{2,3} According to their cell of origin (COO), genetically and phenotypically distinct subgroups of DLBCL (activated B-cell–like, germinal-center B-cell–like [GCB], and unclassified) can be discriminated with different response to treatment and clinical outcomes.⁴⁻⁷

The addition of rituximab (anti-CD20 monoclonal antibody) to the standard cyclophosphamide, doxorubicin, vincristine, and prednisone (CHOP) chemotherapy regimen, given with rituximab either every 2 (R-CHOP14) or 3 (R-CHOP21) weeks, has improved the outcome in patients with DLBCL.^{8,9} However, a relevant portion of patients (30% to 40%) still experience treatment failure.¹⁰ Several systems for predicting prognosis have been developed and the most widely used is the International Prognostic Index (IPI).¹¹ In recent years, modifications of the IPI for DLBCL have shown an improved predictive value,^{12,13} but novel biomarkers are still needed to better identify patients who could benefit from more aggressive therapeutic approaches.

Nowadays, (18)F-fluorodeoxyglucose (18FDG) positron emission tomography/computed tomography (PET/CT) represents the standard for staging and therapeutic response assessment in DLBCL.¹⁴ In fact, the 18FDG-PET/CT scan at the end-of-treatment is the best predictor of DLBCL outcome and its visual evaluation is based on a 5-point scale (Deauville score), where a score of ≤ 3 represents the standard definition of complete remission.^{14,15}

Repeated imaging studies after 1-to-3 R-CHOP cycles (interim positron emission tomography [PET]/computed tomography [CT]) have been proposed to guide treatment decisions, and some studies have suggested that a quantitative approach based on the decrease of the maximum standardized uptake value (Δ SUVmax) between baseline and interim PET/CT may be a more objective tool compared with the visual assessment.¹⁶⁻¹⁹ However, the role of interim PET/CT as a reliable early prognostic factor remains controversial in DLBCL.²⁰⁻²⁵

Several quantitative parameters derived from baseline PET/CT, including the SUVmax, the total lesion glycolysis (TLG), and the metabolic tumor volume (MTV), seem promising biomarkers in various lymphoma subtypes.²⁶⁻³⁰ In particular, some studies reported that an elevated MTV value is a good predictor of poor outcome in patients with DLBCL.³¹⁻³⁵ Improved prognostic models might also be derived from the combination of baseline MTV or TLG and early response on interim PET/CT.^{34,36} In addition, metabolic heterogeneity (MH), a functional PET parameter with known prognostic value in solid tumors,^{37,38} may also have a role in lymphomas.³⁹

Recent research provided preliminary but promising evidence that the integration of quantitative PET parameters with some clinical, biological, and molecular features may allow a more accurate prognostication in DLBCL.^{29,40,41}

The main aim of this study was to test the ability of baseline functional PET parameters in predicting the efficacy of immunochemotherapy in a cohort of DLBCL patients treated with R-CHOP14 in the prospective clinical trial SAKK 38/07. A separate validation cohort of DLBCL patients treated with the standard RCHOP21 was also analyzed.

Patients and methods

All procedures performed in this study were in accordance with the ethical standards of the 1964 Declaration of Helsinki and its later amendments. The study was approved by the Institutional Review Board/Ethics Committee of the participating centers.

Testing set

Among the 156 patients with any stage of untreated DLBCL prospectively enrolled in the SAKK 38/07 trial of the Swiss Group for Clinical Cancer Research,²⁴ 141 with baseline 18FDG-PET/CT scans suitable for imaging postprocessing and a complete clinical follow-up were eligible for this study.

In all cases, the diagnosis of DLBCL was confirmed by central pathology review and the presence of “double-hit” cases with BCL2 and C-MYC rearrangements was ruled out by *in situ* hybridization analyses.

All patients received 6 cycles of R-CHOP (rituximab 375 mg/m², cyclophosphamide 750 mg/m², doxorubicin 50 mg/m², vincristine 1.4 mg/m² on day 1, and prednisone 100 mg/m² for 5 days) every 14 days followed by 2 cycles of rituximab. If indicated by local guidelines, consolidation radiotherapy was allowed and was administered in 21 patients (15%).

PET/CT scans were to be performed in all patients at diagnosis, after 2 cycles of R-CHOP14, and at the end of immunochemotherapy. Treatment details and PET scan protocol have been previously published.²⁴

Validation set

The validation cohort comprised 113 patients with any stage of untreated DLBCL included in a retrospective survey of the PET use in our institutions. Seventy-three patients were obtained from the database of the Oncology Institute of Southern Switzerland, Bellinzona, Switzerland, and 40 from that of the Hematology Unit of the Ospedale Papa Giovanni XXIII, Bergamo, Italy. Age >18 years, 18FDG-PET/CT baseline scan, front-line treatment with 6 cycles of R-CHOP21, and a clinical follow-up of at least 18 months were the criteria of selection. Fifteen patients (13%) received consolidation radiotherapy.

The PET/CT scan was performed baseline, within 2 weeks before the beginning of the treatment, and after the completion of immunochemotherapy.

Baseline whole-body PET/CT was performed with a Biograph 6 HiRez scanner (Siemens, Erlangen, Germany). Acquisition started 60 \pm 5 minutes following the injection of 18FDG (3.5-4 MBq/kg) to the subjects fasting at least 6 hours. PET images (CT corrected for attenuation) were reconstructed with a standard iterative algorithm (3D-OSEM).

Prognostic factors

The IPI,¹¹ the revised-IPI,¹³ and the National Comprehensive Cancer Center Network enhanced-IPI (NCCN-IPI)¹² were calculated for each patient using the clinical information retrieved from either the SAKK38/07 study database or the hospital records, as appropriate. The COO was centrally determined by immunohistochemistry using the Hans algorithm.⁴²

PET/CT images analysis

For the present study, all the PET/CT scans were centrally evaluated following a standard protocol with dedicated imaging software (MM Oncology, Syngo.via, Siemens). The lymphoma lesions were segmented using an algorithm with a fixed threshold at 2.5 SUV value⁴³ for MTV estimation, then SUVmax and TLG were calculated automatically. MH of the target lesion (ie, the lesion with

Table 1. Comparison of baseline patient characteristics and functional PET parameters in the testing (N = 141) and validation (N = 113) cohorts

Clinical features	Testing set (SAKK 38/07), n (%)	Validation set, n (%)	P*
Sex			.833
Male	73 (52)	57 (51)	
Female	68 (48)	56 (49)	
Age			
Median; IQR	59 y; 49-68	70 y; 57-76	<.0001
>60 y	68 (48)	80 (71)	<.0001
Ann Arbor stage			.275
1	15 (11)	15 (13)	
2	46 (33)	34 (30)	
3	33 (23)	17 (15)	
4	47 (33)	47 (42)	
Bulky disease			
>7.5 cm	70 (50)	50 (44)	.392
>10 cm	42(30)	29 (26)	.375
Elevated LDH	68 (48)	56 (50)	.726
ECOG PS >1	10 (7)	9 (8)	.793
Extranodal sites >1	34 (24)	25 (22)	.709
COO (Hans algorithm)	N = 113	N = 99	<.001
GCB	29 (26)	58 (59)	
Non-GCB	84 (74)	41 (41)	
IPI group			.085
Low risk	65 (46)	36 (32)	
Intermediate-low risk	32 (23)	39 (34)	
Intermediate-high risk	27 (19)	25 (22)	
High risk	17 (12)	13 (12)	
R-IPI group			.111
Low risk	22 (16)	8 (7)	
Intermediate risk	75 (53)	67 (59)	
High risk	44 (31)	38 (34)	
NCCN-IPI group			.007
Low risk	22 (16)	7 (6)	
Intermediate-low risk	67 (47)	42 (37)	
Intermediate-high risk	39 (28)	51 (45)	
High risk	13 (9)	13 (12)	
PET parameters, median (IQR)			
SUVmax	19.9 (15.1-28.2)	21 (13.5–28.1)	.95
MTV	386 (143-1119)	342 (78- 822)	.12
TLG	2618 (824-8906)	2325 (472-5390)	.08
Metabolic heterogeneity (AUC-CSH)	0.45 (0.40-0.50)	0.49 (0.46-0.53)	<.001

Bold P values indicate statistically significant results ($P < .05$).

ECOG, Eastern Cooperative Oncology Group; PS, performance status.

* χ^2 test for comparison of frequencies and Mann-Whitney U test for comparison of medians.

the highest 18FDG uptake) was measured in each patient using the area under curve of cumulative SUV-volume histogram (AUC-CSH) method as previously published.³⁹

Statistics

Quantitative variables were expressed by medians and interquartile ranges (IQR). Medians were compared using the Mann-Whitney U test. Differences between the frequencies of categorical data were assessed with the χ^2 test. The PET-associated functional continuous parameters were analyzed as dichotomized variables, using receiver-operating characteristic (ROC) analysis to identify the optimal cutoff point to discriminate subgroups with different progression-free survival (PFS) and overall survival (OS). Survival functions were defined according to the revised National Cancer Institute criteria⁴⁴ and estimated by the Kaplan-Meier method, and patient groups were compared by the log-rank test. Cox proportional hazard models were used for multivariable analysis and the estimation of hazard ratios (HRs). The predictive accuracy of the different prognostic indices in the identification of patients at risk of shorter survival was compared using the Harrell C concordance probability estimate (CPE),⁴⁵ which calculates the probability of agreement for any pair of patients in which agreement means that the patient with the shorter survival time also has the higher predicted risk. Because C statistics range from 0.5 (random concordance) to 1 (perfect concordance), higher values of C indicate better discrimination. Their relative quality was further assessed using an in-sample fit approach to model selection (Akaike information criterion [AIC]),⁴⁶ which estimates the likelihood of a prognostic model to predict future outcomes. The best model is the one that, in comparison with all the others, has the minimum AIC (ie, best fit). $P < .05$ was considered statistically significant. Negative predictive value (NPV) and positive predictive value (PPV) were calculated according to standard definitions. Both testing and validation datasets were explored using a recursive-partitioning classification tree method (implemented into the ctree function of the R package party) to develop unbiased prognostic models based on dichotomized variables.⁴⁷ This procedure enables the hierarchical classification of the prognostic covariates, from the most important, which splits the primary node (entire population), to those which extend to the terminal nodes (risk groups).

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, as appropriate.

Results

Testing set (SAKK38/07)

Detailed clinical features and outcome of the patients enrolled in the SAKK 38/07 study have been published previously.²⁴ Table 1 summarizes the demographic and clinical characteristics of the 141 patients included in the present analysis and reports the estimation of their functional PET parameters at baseline. At a median follow-up of 64 months (IQR, 60-67 months), 30 progressions of disease and 23 deaths were recorded. The estimated 5-year PFS and OS rates were 76% (95% confidence interval [CI], 67-82) and 84% (95% CI, 77-90), respectively.

Univariable analysis

Among the clinical features evaluable at presentation, no impact on outcome appeared related to sex, age (>60 years), serum lactate

Table 2. Results of the ROC analysis used to identify optimal cutoff points for PFS and OS in the testing (N = 141) and validation (N = 113) cohorts

Functional PET parameters	ROC analysis for PFS					ROC analysis for OS					
	Cutoff point	Sensitivity, %	Specificity, %	AUC	P	Cutoff point	Sensitivity, %	Specificity, %	AUC	P	
Testing set	SUVmax	20	60	55	0.556	.342	24.2	74	40	0.550	.433
	MTV, mL	931	57	75	0.629	.029	1149	52	81	0.670	.011
	TLG	3960	63	62	0.620	.034	6991	57	73	0.661	.012
	MH (AUC-CSH)	0.43	53	66	0.545	.474	0.43	52	69	0.555	.431
Validation set	SUVmax	31	27	86	0.507	.917	31	32	86	0.582	.267
	MTV, mL	336	73	54	0.629	.029	336	79	54	0.637	.070
	TLG	3186	55	66	0.574	.318	3574	63	70	0.634	.079
	MH (AUC-CSH)	0.47	53	66	0.527	.730	0.46	42	85	0.648	.046

Bold P values indicate statistically significant results ($P < .05$).

dehydrogenase (LDH), stage, extranodal disease, and bulky disease (>7.5 cm). An impaired performance status (Eastern Cooperative Oncology Group score > 1) was associated with a shorter OS ($P = .0342$), but did not show a significant effect on PFS. The IPI, R-IPI, and NCCN-IPI were also associated with OS (trend test $P = .0442$, $.0279$, and $.0045$, respectively) but had only a borderline impact on PFS (trend test P , $.0590$, $.0498$, and $.0516$, respectively). Table 2

reports the optimal cutoff points of the dichotomized PET parameters estimated by ROC analysis for PFS and OS. In the whole cohort, elevated MTV and TLG were significantly associated with worse PFS and OS, whereas SUVmax and MH had no effect on outcomes (Table 3). However, when the analysis was limited to the 126 patients with stages 2-4, MH became significantly associated with PFS ($P = .03$) and had a borderline effect also on OS ($P = .07$). Moreover,

Table 3. Univariable analysis of PFS and OS by dichotomized (below vs above the ROC cutoff point) functional PET parameters in the validation and testing sets

Functional PET parameters	Univariable analysis of PFS					Univariable analysis of OS					
	5-y PFS (95% CI), %	Log-rank P	HR (95%CI)	NPV	PPV	5-y OS (95% CI), %	Log-rank P	HR (95%CI)	NPV	PPV	
Testing set	SUVmax		.164	0.6 (0.3-1.2)	84	27		.196	0.5 (0.2-1.4)	89	19
	Low	72 (58-81)						81 (71-88)			
	high	80 (67-88)						90 (78-96)			
	MTV (mL)		.0005	3.4 (1.6-7.0)	86	40		.0001	4.4 (1.9-10)	90	35
	Low	83 (72-90)						91 (83-95)			
	High	61 (44-74)						65 (46-78)			
	TLG		.0078	2.6 (1.3-5.6)	86	31		.0021	3.4 (1.5-7.8)	90	29
	Low	83 (72-90)						90 (81-95)			
	High	65 (50-77)						73 (57-84)			
	MH		.0780	1.9 (0.9-3.9)	84	29		.1011	2.0 (0.9-4.4)	87	22
	Low	80 (69-88)						88 (79-93)			
	High	69 (54-80)						79 (65-88)			
Validation set	SUVmax		.0935	2.2 (0.9-5.6)	83	32		.0432	2.6 (0.99-6.9)	86	32
	Low	77 (63-86)						79 (63-89)			
	High	67 (41-84)						63 (34-82)			
	MTV (mL)		.021	2.9 (1.1-7.3)	89	28		.0090	3.9 (1.3-11.8)	93	26
	Low	85 (67-93)						92 (80-97)			
	High	66 (48-79)						62 (41-77)			
	TLG		.0632	2.2 (0.9-5.0)	86	27		.0107	3.1 (1.2-7.8)	89	27
	Low	82 (67-90)						83 (65-92)			
	High	64 (41-80)						63 (39-79)			
	MH		.1547	1.8 (0.8-4.2)	84	28		.0433	2.5 (0.99-6.0)	88	28
	Low	80 (64-89)						85 (74-92)			
	High	67 (47-81)						56 (26-78)			

Bold P values indicate statistically significant results ($P < .05$).

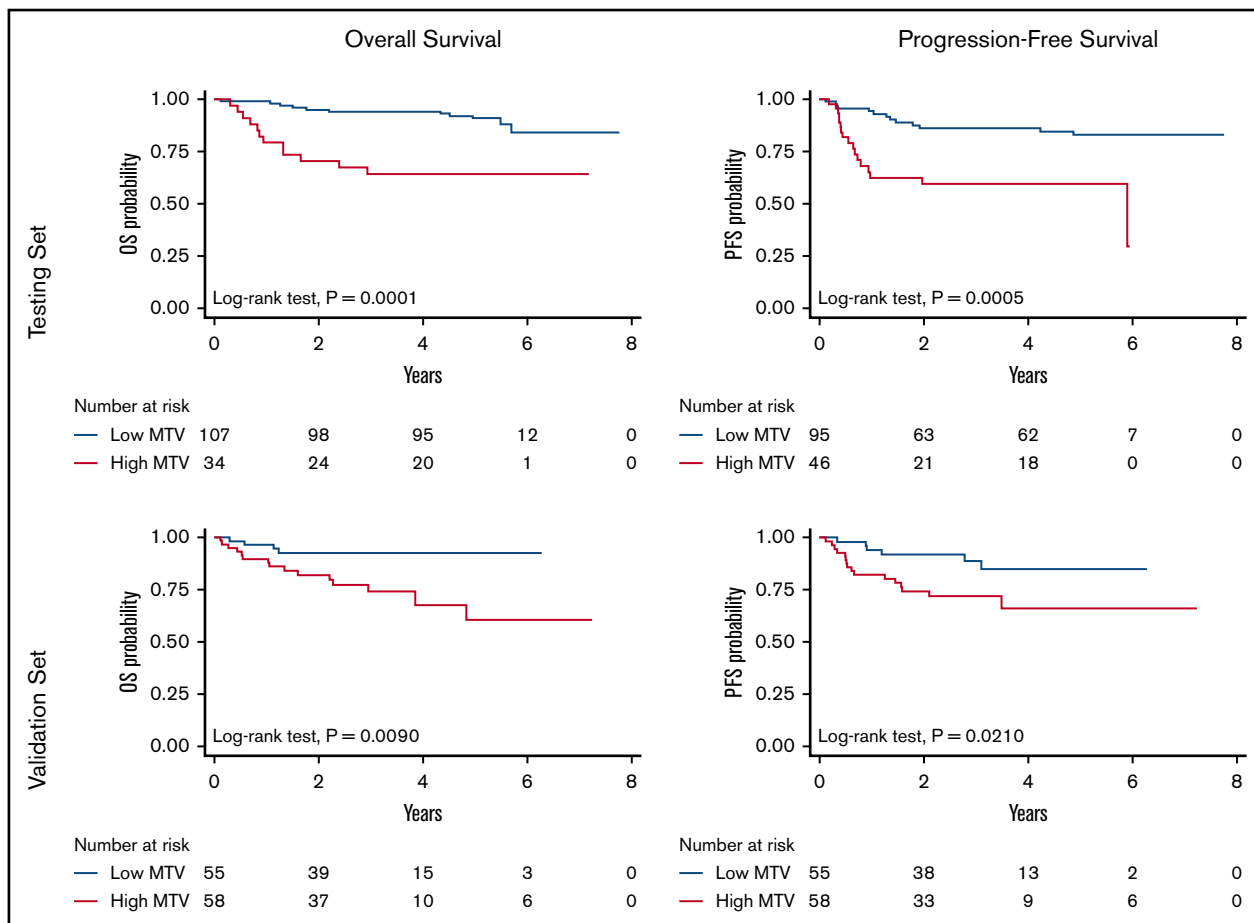


Figure 1. OS and PFS according to the MTV in the testing and in the validation cohorts.

elevated MH was associated with shorter PFS ($P = .02$) and OS ($P = .009$) in the subset of patients with elevated MTV. Among the dichotomized IPI factors, at univariable analysis, a significant association was found between baseline MTV and serum LDH, extranodal disease and stage, not with age or performance status.

PFS at 5 years was 83% (95% CI, 72-90) for patients with low MTV vs 61% (95% CI, 44-74) for those with high MTV (log-rank test, $P = .0005$), whereas OS was 91% (95% CI, 83-95) vs 65% (95% CI, 46-78) (log-rank test, $P = .0001$), respectively (Figure 1, top). Outcome at 5 years was anticipated by dichotomized MTV with very high NPV (86% and 90% for PFS and OS, respectively) but lower PPV (40% for PFS and 35% for OS, respectively).

Integrated MTV-MH model

Because of the MH effect on the outcome of patients with high MTV, we explored the prognostic impact of a model integrating MTV and MH (Figure 2, top). The combination of these 2 parameters improved the performance of MTV alone to predict treatment outcome (log-rank test for trend, $P < .001$ for both PFS and OS). Among the patients with high MTV, those with high MH showed a significantly higher risk of progression (HR, 5.6; 95% CI, 1.8-17) and death (HR, 9.5; 95% CI, 1.7-52). The PPV of this prognostic model for the definition of the poorer risk patients was 57% for PFS and 62% for OS, respectively. Notably, an unsupervised approach with a classification tree method based on binary recursive

partitioning (which included all the functional PET parameters and the dichotomized variables that contribute to the international prognostic indices) also found MTV and HM to be the most important factors, leading to a superimposable predictive model for both OS and PFS (Figure 3, top).

Multivariable analysis

At multivariable analysis (stepwise Cox regression including the functional PET parameters, which were significant at univariable analysis) only MTV retained statistical significance for both OS (HR, 4.4; 95% CI, 1.9-10; $P < .01$) and PFS (HR, 3.4; 95% CI, 1.6-7.0; $P = .01$). The prognostic impact of MTV on OS and PFS did not change after controlling for the performance status and for the IPI, R-IPI, and NCCN-IPI.

The model integrating MTV and MH was also the only parameter retaining a significant prognostic effect on OS (HR, 3.1; 95% CI, 1.9-5.0; $P < .001$) and PFS (HR, 2.4; 95% CI, 1.6-3.6; $P < .001$) in a stepwise Cox regression including the IPIs.

Validation set

The validation set included patients with significantly older age and showed a lower proportion of patients with good risk according to the NCCN-IPI, but the distribution of risk groups according to either IPI or R-IPI was not statistically different (Table 1). The median follow-up (36 months; IQR, 28-50) was shorter in comparison with

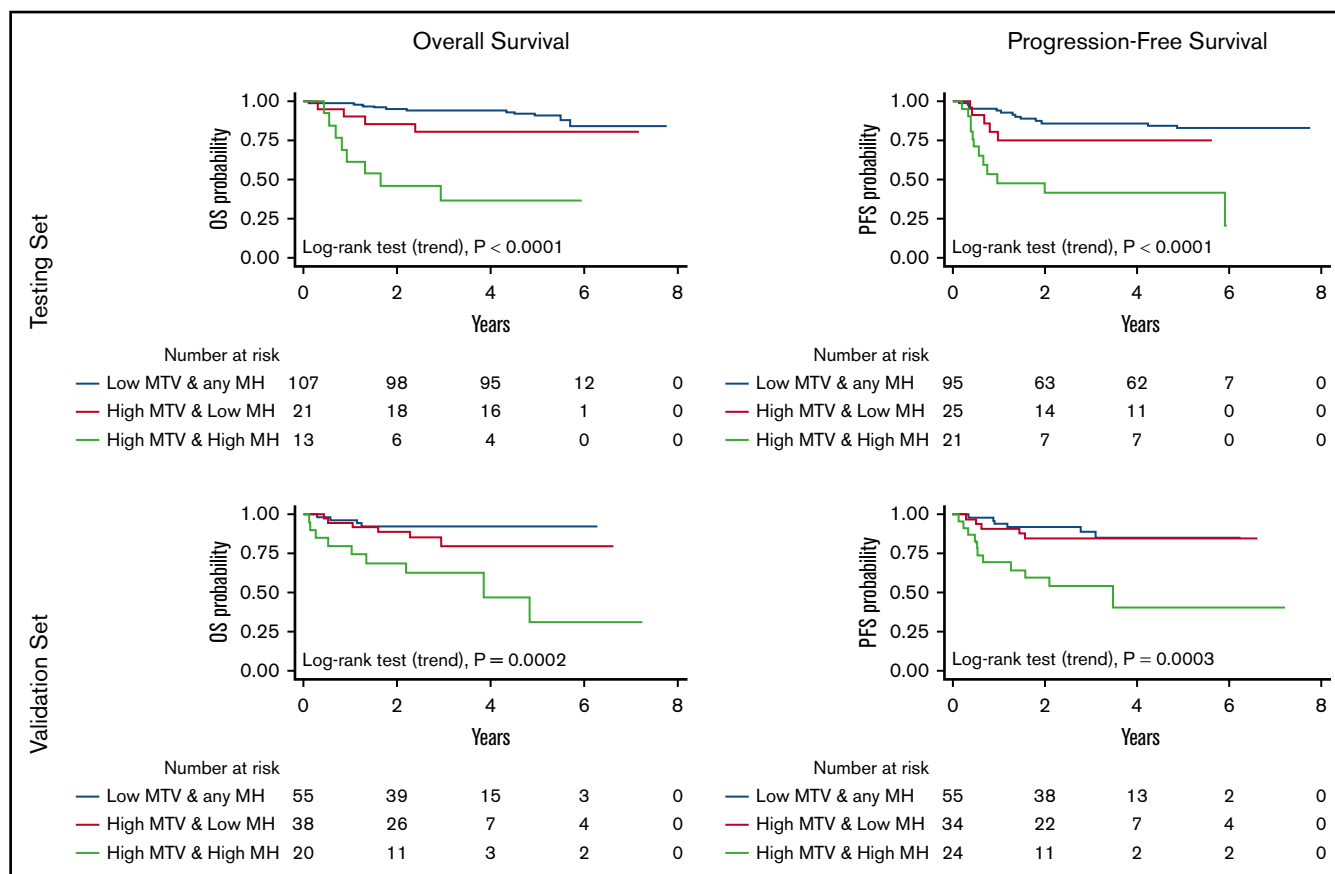


Figure 2. OS and PFS according to the model built integrating the MTV and MH either in the testing or in the validation cohort.

the SAKK 38/07 cohort ($P < .001$). Twenty-two patients had disease progressions and 19 died. The estimated 5-year PFS and OS rates were 75% (95% CI, 63-83) and 76% (95% CI, 63-85), respectively. All the dichotomized functional PET parameters at baseline had a significant impact on OS at univariable analysis. However, only MTV showed a statistically significant association with PFS (Table 3; Figure 1, bottom). Analogous to the testing set, low values of MTV accurately identified the subset of patients with good treatment outcome (NPV, 89% and 93% for PFS and OS, respectively). However, high MTV values were less efficient in the discrimination of the poor prognosis patients (PPV, 28% and 26% for PFS and OS, respectively). Also, in keeping with the testing set, an impaired performance status predicted OS ($P = .0023$), but not PFS. IPI and NCCN-IPI were also associated with OS (trend test $P = .0465$ and 0.0261 , respectively) but had no effect on PFS; R-IPI did not affect the OS nor PFS. At multivariable analysis (stepwise Cox regression controlling for the performance status and the IPIs) an elevated MTV value remained independently associated with a shorter PFS (HR, 2.9; 95% CI, 1.1-7.3; $P = .028$) and OS (HR, 3.6; 95% CI, 1.2-11; $P = .023$).

Elevated MH was associated with shorter OS ($P = .043$) in the entire cohort and with both PFS ($P = .006$) and OS ($P = .016$) in the subset of patients with elevated MTV. The model integrating MTV and MH showed a PPV of 46% for PFS and 45% for OS and confirmed in the validation set its ability to improve upon MTV alone the identification of the poor-risk patients with shorter PFS (HR, 5.6;

95% CI, 1.7-18) and OS (HR, 7.6; 95% CI, 2.0-28) (Figure 2, bottom). At multivariable analysis, the prognostic efficacy of this model for PFS (HR, 2.5; 95% CI, 1.5-4.3; $P = .001$) and OS (HR, 2.8; 95% CI, 1.6-5.1; $P = .001$) was maintained after controlling for the IPIs. Analogous to the testing set, the regression tree analysis generated again a predictive model based on MTV and MH, for either OS or PFS (Figure 3, bottom).

Comparison of PET-derived models and prognostic indices

The discriminatory power and predictive accuracy of this PET-derived model appeared superior, in both the testing and the validation cohorts, to the ones of the international clinical indices (Table 4) when assessed by using either the AIC⁴⁶ or a CPE estimator (Harrell C statistic).⁴⁵

Discussion

Despite the potential power limitation because of the overall good outcome of patients in our cohorts, resulting in a relatively low number of events, this study further substantiates the high predictive value of baseline MTV in DLBCL³¹⁻³⁵ and provides a first evidence that MH combined with other functional volumetric PET parameters may allow to foresee treatment outcomes in DLBCL. The main strengths of this study are the uniform treatment, the centralized analysis of PET scans, and, above all, the reproducibility assessment in an independent validation set of patients of the observations made in the cohort

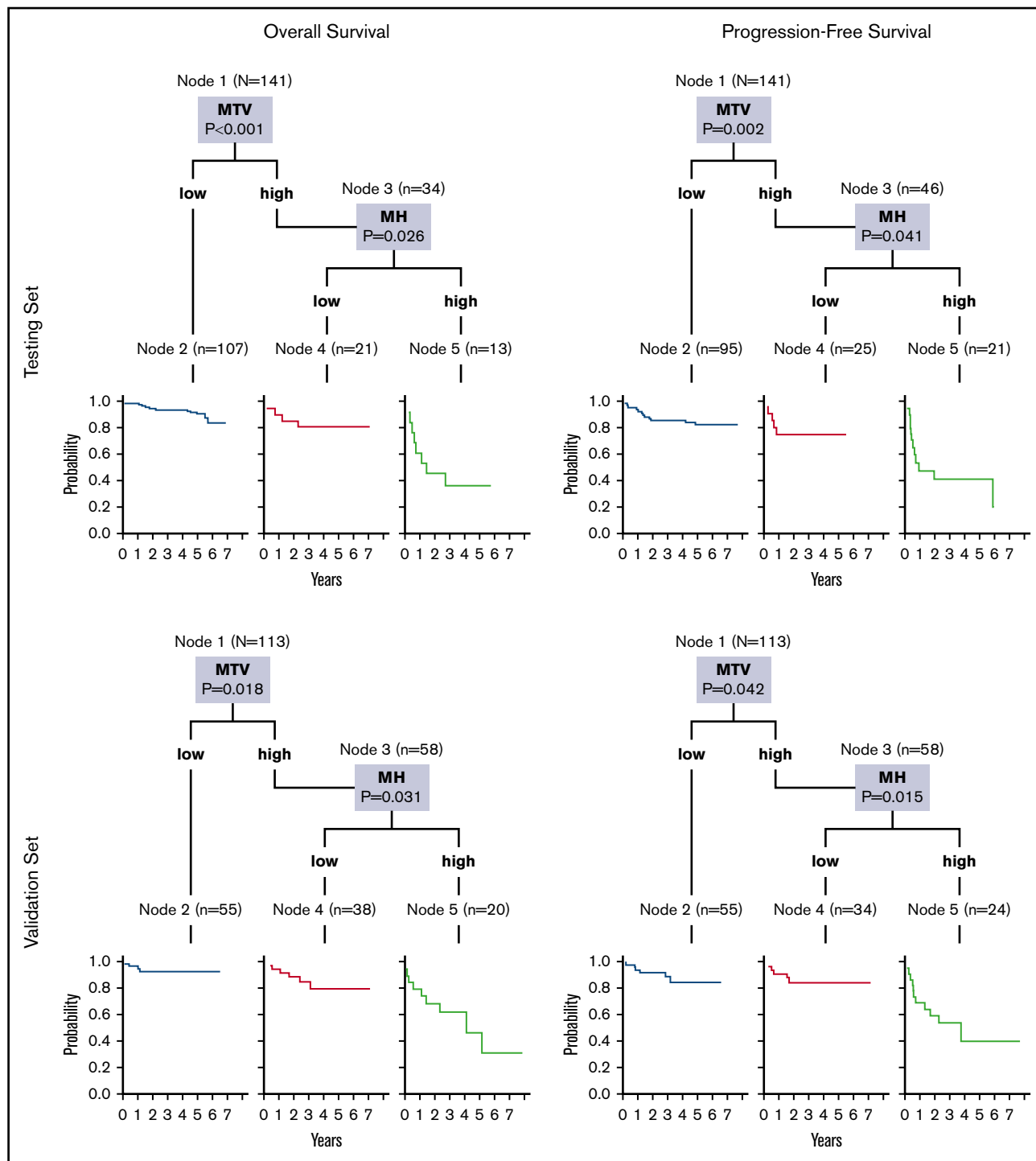


Figure 3. The classification trees showing the prognostic signature consisting of 2 variables, MTV and MH, whose dichotomization was significantly correlated with survival. MTV alone, when its value is low, identifies the prognostic category node 2 with best survival probability. To separate the prognostic categories node 4, with intermediate survival rates, from node 5, with poorer survival, the value of MTV must be high (node 3), and that of MH must be low or high, respectively. The Kaplan-Meier curves show statistically different overall and progression-free survival, either in both the testing and the validation patient sets. For each inner node, the Bonferroni-adjusted *P* values are given, which take into account the multiple testing and may differ from the *P* values of the log-rank test in the univariable analysis (Table 3).

of patients enrolled in the SAKK38/07 study. Differences (eg, in age, COO, and MH distribution) between the testing and validation sets support the potential general validity of the proposed model, too.

Because taking into account the interaction of individual factors in determining an outcome endpoint (such as PFS or OS) can be difficult, their selection in a prognostic model is typically complex.

Table 4. Comparison of clinical indices and models based on MTV and MH

	Prognostic model	PFS		OS	
		AIC	CPE	AIC	CPE
Testing set	MTV dichotomized	265	0.6436 (± 0.0433)	207	0.6847 (± 0.0440)
	MTV+MH model	260	0.6684 (± 0.0440)	201	0.7065 (± 0.0443)
	IPI	273	0.5908 (± 0.0513)	215	0.6086 (0.0584)
	R-IPI	272	0.5908 (± 0.0498)	214	0.6175 (± 0.0564)
	NCCN-IPI	272	0.5857 (± 0.0511)	211	0.6409 (± 0.0578)
Validation set	MTV dichotomized	192	0.6356 (± 0.0556)	162	0.6346 (± 0.0607)
	MTV+MH model	186	0.6944 (± 0.0586)	156	0.6884 (± 0.0638)
	IPI	194	0.6143 (± 0.0613)	165	0.6171 (± 0.0668)
	R-IPI	196	0.5586 (± 0.0556)	166	0.5933 (± 0.0606)
	NCCN-IPI	196	0.5729 (± 0.0592)	164	0.6150 (± 0.646)

We built the model that combines MTV and MH on the basis of the impact of MH in the subsets of patients with high MTV or nonlocalized disease. This model is also in keeping with a recent meeting abstract⁴⁸ and with the observations we previously made in a primary mediastinal lymphoma study, where the subset of patients with elevated MH and TLG values had a particularly dismal prognosis.³⁹ The validity of the model was then confirmed by an unsupervised approach using a methodology (regression tree analysis based on unbiased binary recursive partitioning) that is widely used in data mining.⁴⁷ This alternative statistical approach maximizes sensitivity by identifying patients truly at risk, minimizes misclassification of low-risk patients, and can stratify patients into different levels of risk.⁴⁷ The regression tree analysis in our datasets ended up with the same 2 PET-derived factors (MTV and MH), producing a simple decision tree that appears accurate and easy to apply.

The present study provides the first demonstration that MH can play a critical role in DLBCL, where (especially in the presence of large MTVs) it appears to be associated with treatment resistance and treatment failure. In both the testing and the validation cohorts, MH effectively discriminated, among patients with increased MTV, those with truly poorer prognosis.

MH describes a complex phenomenon (depending on cell metabolism, proliferation, blood flow, and hypoxia) that reflects the clinical and molecular heterogeneity of the disease.³⁹ It appears plausible that MH introduced into our prognostic model additional biological information that contributed to further characterize the prognostic risk defined by the simple metrics of the tumor burden provided by the MTV estimation.

Main limitations of the study, which may make the use of MTV and MH in routine clinical practice still premature, are the absence of standardized methodology and the lack of agreement on the optimal estimation of volume-based PET parameters. Different procedures and a wide range of threshold levels have been proposed to calculate MTV.^{35,43,49} Moreover, the prognostic effect of MTV has been estimated using cut-points that are heavily cohort-dependent, being generated by ROC analysis of small series. In this study, we chose to use a segmentation method with a fixed-threshold at $SUV = 2.5$ ⁴³ that, compared with the widely used percentage threshold at 41% of SUV_{max} , maintains a similar accuracy and offers better tumor coverage,⁴¹ particularly in lesions with heterogeneous FDG uptake distribution.⁴⁹

Although the lack of standardization endangers the reproducibility of the results obtained in retrospective studies, there is increasing evidence of the prognostic value of quantitative parameters obtained from 18FDG-PET/CT in patients with different non-Hodgkin lymphoma subtypes.^{17,26,28,50,51} We sought to use the quantitative parameters from PET scans to discriminate patient with different outcomes. The PET-based prognostic models appeared superior to the clinical and pathological prognostic indices that are widely and routinely used. Prior reports have already indicated that volumetric PET parameters may be better outcome predictors than the IPI.^{31,52} Only 1 study, which analyzed a limited number of quantitative whole-body 18FDG-PET/CT and failed to show any independent prognostic utility of PET parameters, suggested that the NCCN-IPI may be superior⁵³; however, the latter study adopted a dichotomized NCCN index which is not currently used.

Work is in progress to solve the previously mentioned methodological problems⁵⁴⁻⁵⁶ and the integration of the functional PET parameters with clinical indices and molecular features is expected to become an important tool for a more precise identification of the very-high-risk group of patients and for the subsequent development of patient-tailored treatment approaches.^{29,56,57} In this perspective, our results indicate that models based on the combination of functional parameters from baseline 18FDG-PET/CT may represent a simple and powerful tool to help early identification of patients with very high risk for initial treatment failure, who can be selected for more intensive treatment (eg, stem cell transplant or CAR-T cell strategies). If confirmed in other cohorts, these results could influence the design of future clinical trials and may offer an opportunity to reduce the number of patients dying from refractory DLBCL.

Acknowledgments

The authors thank the investigators, research nurses, and data managers of the SAKK 38/07 trial at each study center, as well as the central study team at the SAKK coordinating center for their administrative support in data collection and study conduction. The authors also thank their colleagues Franco Cavalli and Beat Thürlimann for critically reading the manuscript and for their fruitful feedback and thoughtful discussion; special thanks go to Sarah Jane Ortelli Giannakis for her editorial support and to Rita Gianascio Gianocca for the secretarial help. The authors express gratitude to

Francesca Ceriani, MSMBE, for her assistance in the estimation of metabolic heterogeneity.

The SAKK 38/07 trial was subsidized by a grant from Amgen (Switzerland) and a grant (OCS 02270-08-2008) from Oncosuisse (Switzerland). The present analysis was partly supported by a grant (ABREOC 22008-262) from the Ente Ospedaliero Cantonale, Bellinzona (Switzerland). The SAKK organization is supported by the Swiss State Secretariat for Education, Research and Innovation (SERI).

Authorship

Contribution: L. Ceriani and E.Z. designed the study, performed research, analyzed the data, and wrote the paper; L. Cascione contributed to the study design data analysis and manuscript writing; S.H. and S.S. reviewed the statistical analysis; and all authors contributed to data collection, reviewed and approved the manuscript, and shared final responsibility for the decision to submit.

The members of the Swiss Group for Clinical Cancer Research (SAKK) appear in the Appendix.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

ORCID profiles: Luca Ceriani, 0000-0002-6371-097X; G.G., 0000-0002-6789-5123; L. Cascione, 0000-0002-4606-0637; L.G., 0000-0003-0230-0974; S.H., 0000-0002-2265-6559; A.R., 0000-0002-3739-7502; C.M., 0000-0002-8838-6360; E.Z., 0000-0002-5522-6109.

Correspondence: Emanuele Zucca, Medical Oncology Clinic, Oncology Institute of Southern Switzerland, Ospedale San Giovanni, CH-6500 Bellinzona, Switzerland; e-mail: emanuele.zucca@eoc.ch.

Appendix: study group members

Members of the Swiss Group for Clinical Cancer Research (SAKK) include: Mario Bargetzi, Francesco Bertoni, Roger Burkhard, Clemens Caspar, L. Ceriani, S.D., Christoph Driessen, Natalie Fischer, Thomas Hany, Felicitas Hitz, Lorenz M. Jost, Dieter Köberle, Fatime Krasniqi, Nicolas Mach, C.M., G.M., Ulrich Mey, Walter Mingrone, A.A.M., Urban Novak, Thomas Pabst, Miklos Pless, Christoph Renner, Andrei Samarin, Mathias Schmid, A.S., Georg Tscherry, Michèle Voegeli, Reinhard Zenhäusern, Thilo Zender, Thorsten Zenz, and E.Z.

References

1. Armitage JO, Gascoyne RD, Lunning MA, Cavalli F. Non-Hodgkin lymphoma. *Lancet*. 2017;390(10091):298-310.
2. Chapuy B, Stewart C, Dunford AJ, et al. Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes [published correction appears in *Nat Med*. 2018;24(8):1290-1291 and *Nat Med*. 2018;24(8):1292]. *Nat Med*. 2018;24(5):679-690.
3. Schmitz R, Wright GW, Huang DW, et al. Genetics and pathogenesis of diffuse large B-cell lymphoma. *N Engl J Med*. 2018;378(15):1396-1407.
4. Meyer PN, Fu K, Greiner TC, et al. Immunohistochemical methods for predicting cell of origin and survival in patients with diffuse large B-cell lymphoma treated with rituximab. *J Clin Oncol*. 2011;29(2):200-207.
5. Lenz G, Wright G, Dave SS, et al; Lymphoma/Leukemia Molecular Profiling Project. Stromal gene signatures in large-B-cell lymphomas. *N Engl J Med*. 2008;359(22):2313-2323.
6. Thieblemont C, Briere J, Mounier N, et al. The germinal center/activated B-cell subclassification has a prognostic impact for response to salvage therapy in relapsed/refractory diffuse large B-cell lymphoma: a bio-CORAL study. *J Clin Oncol*. 2011;29(31):4079-4087.
7. Wilson WH, Young RM, Schmitz R, et al. Targeting B cell receptor signaling with ibrutinib in diffuse large B cell lymphoma. *Nat Med*. 2015;21(8):922-926.
8. Delarue R, Tilly H, Mounier N, et al. Dose-dense rituximab-CHOP compared with standard rituximab-CHOP in elderly patients with diffuse large B-cell lymphoma (the LNH03-6B study): a randomised phase 3 trial. *Lancet Oncol*. 2013;14(6):525-533.
9. Cunningham D, Hawkes EA, Jack A, et al. Rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisolone in patients with newly diagnosed diffuse large B-cell non-Hodgkin lymphoma: a phase 3 comparison of dose intensification with 14-day versus 21-day cycles. *Lancet*. 2013;381(9880):1817-1826.
10. Gisselbrecht C, Glass B, Mounier N, et al. Salvage regimens with autologous transplantation for relapsed large B-cell lymphoma in the rituximab era. *J Clin Oncol*. 2010;28(27):4184-4190.
11. International Non-Hodgkin's Lymphoma Prognostic Factors Project. A predictive model for aggressive non-Hodgkin's lymphoma. *N Engl J Med*. 1993;329(14):987-994.
12. Zhou Z, Sehn LH, Rademaker AW, et al. An enhanced International Prognostic Index (NCCN-IPI) for patients with diffuse large B-cell lymphoma treated in the rituximab era. *Blood*. 2014;123(6):837-842.
13. Sehn LH, Berry B, Chhanabhai M, et al. The revised International Prognostic Index (R-IPI) is a better predictor of outcome than the standard IPI for patients with diffuse large B-cell lymphoma treated with R-CHOP. *Blood*. 2007;109(5):1857-1861.
14. Cheson BD, Fisher RI, Barrington SF, et al; United Kingdom National Cancer Research Institute. Recommendations for initial evaluation, staging, and response assessment of Hodgkin and non-Hodgkin lymphoma: the Lugano classification. *J Clin Oncol*. 2014;32(27):3059-3068.
15. Barrington SF, Mikhaeel NG, Kostakoglu L, et al. Role of imaging in the staging and response assessment of lymphoma: consensus of the International Conference on Malignant Lymphomas Imaging Working Group. *J Clin Oncol*. 2014;32(27):3048-3058.
16. Itti E, Lin C, Dupuis J, 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(4):527-533.

17. Casasnovas RO, Meignan M, Berriolo-Riedinger A, et al; Groupe d'étude des lymphomes de l'adulte (GELA). SUVmax reduction improves early prognosis value of interim positron emission tomography scans in diffuse large B-cell lymphoma. *Blood*. 2011;118(1):37-43.
18. Safar V, Dupuis J, Itti E, et al. Interim [18F]fluorodeoxyglucose positron emission tomography scan in diffuse large B-cell lymphoma treated with anthracycline-based chemotherapy plus rituximab. *J Clin Oncol*. 2012;30(2):184-190.
19. Itti E, Meignan M, Berriolo-Riedinger A, 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 Δ SUVmax. *Eur J Nucl Med Mol Imaging*. 2013;40(9):1312-1320.
20. Moskowitz CH, Schöder H, Teruya-Feldstein J, et al. Risk-adapted dose-dense immunochemotherapy determined by interim FDG-PET in advanced-stage diffuse large B-cell lymphoma. *J Clin Oncol*. 2010;28(11):1896-1903.
21. Cox MC, Ambrogi V, Lanni V, et al. Use of interim [18F]fluorodeoxyglucose-positron emission tomography is not justified in diffuse large B-cell lymphoma during first-line immunochemotherapy. *Leuk Lymphoma*. 2012;53(2):263-269.
22. Dührsen U, Müller S, Hertenstein B, et al; PETAL Trial Investigators. Positron Emission Tomography-Guided Therapy of Aggressive Non-Hodgkin Lymphomas (PETAL): a multicenter, randomized phase III trial. *J Clin Oncol*. 2018;36(20):2024-2034.
23. Gisselbrecht C. Positron emission tomography-guided therapy of aggressive non-Hodgkin lymphoma: standard of care after the PETAL study? *J Clin Oncol*. 2018;JCO1800498.
24. Mamot C, Klingbiel D, Hitz F, et al. Final results of a prospective evaluation of the predictive value of interim positron emission tomography in patients with diffuse large B-cell lymphoma treated with R-CHOP-14 (SAKK 38/07). *J Clin Oncol*. 2015;33(23):2523-2529.
25. Pregno P, Chiappella A, Bellò M, et al. Interim 18-FDG-PET/CT failed to predict the outcome in diffuse large B-cell lymphoma patients treated at the diagnosis with rituximab-CHOP. *Blood*. 2012;119(9):2066-2073.
26. Kim CY, Hong CM, Kim DH, et al. Prognostic value of whole-body metabolic tumour volume and total lesion glycolysis measured on ¹⁸F-FDG PET/CT in patients with extranodal NK/T-cell lymphoma. *Eur J Nucl Med Mol Imaging*. 2013;40(9):1321-1329.
27. Kanoun S, Rossi C, Berriolo-Riedinger A, et al. Baseline metabolic tumour volume is an independent prognostic factor in Hodgkin lymphoma. *Eur J Nucl Med Mol Imaging*. 2014;41(9):1735-1743.
28. 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(8):950-956.
29. Cottreau AS, Lanic H, Mareschal S, et al. Molecular profile and FDG-PET/CT total metabolic tumor volume improve risk classification at diagnosis for patients with diffuse large B-cell lymphoma. *Clin Cancer Res*. 2016;22(15):3801-3809.
30. Zhou M, Chen Y, Huang H, Zhou X, Liu J, Huang G. Prognostic value of total lesion glycolysis of baseline 18F-fluorodeoxyglucose positron emission tomography/computed tomography in diffuse large B-cell lymphoma. *Oncotarget*. 2016;7(50):83544-83553.
31. Sasanelli M, Meignan M, Haioun C, 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(11):2017-2022.
32. Chang CC, Cho SF, Chuang YW, et al. Prognostic significance of total metabolic tumor volume on ¹⁸F-fluorodeoxyglucose positron emission tomography/computed tomography in patients with diffuse large B-cell lymphoma receiving rituximab-containing chemotherapy. *Oncotarget*. 2017;8(59):99587-99600.
33. Kim J, Hong J, Kim SG, et al. Prognostic value of metabolic tumor volume estimated by (18)F-FDG positron emission tomography/computed tomography in patients with diffuse large B-cell lymphoma of stage II or III disease. *Nucl Med Mol Imaging*. 2014;48(3):187-195.
34. Mikhaeel NG, Smith D, Dunn JT, et al. Combination of baseline metabolic tumour volume and early response on PET/CT improves progression-free survival prediction in DLBCL. *Eur J Nucl Med Mol Imaging*. 2016;43(7):1209-1219.
35. Schmitz C, Hüttmann A, Müller SP, et al. Dynamic risk assessment based on positron emission tomography scanning in diffuse large B-cell lymphoma: post-hoc analysis from the PETAL trial. *Eur J Cancer*. 2020;124:25-36.
36. Zhang YY, Song L, Zhao MX, Hu K. A better prediction of progression-free survival in diffuse large B-cell lymphoma by a prognostic model consisting of baseline TLG and $\% \Delta$ SUV_{max}. *Cancer Med*. 2019;8(11):5137-5147.
37. Lee M, Lee H, Cheon GJ, et al. Prognostic value of preoperative intratumoral FDG uptake heterogeneity in patients with epithelial ovarian cancer. *Eur Radiol*. 2017;27(1):16-23.
38. Eary JF, O'Sullivan F, O'Sullivan J, Conrad EU. Spatial heterogeneity in sarcoma 18F-FDG uptake as a predictor of patient outcome. *J Nucl Med*. 2008;49(12):1973-1979.
39. Ceriani L, Milan L, Martelli M, et al. Metabolic heterogeneity on baseline 18FDG-PET/CT scan is a predictor of outcome in primary mediastinal B-cell lymphoma. *Blood*. 2018;132(2):179-186.
40. Toledano MN, Desbordes P, Banjar A, et al. Combination of baseline FDG PET/CT total metabolic tumour volume and gene expression profile have a robust predictive value in patients with diffuse large B-cell lymphoma. *Eur J Nucl Med Mol Imaging*. 2018;45(5):680-688.
41. Shagera OA, Cheon GJ, Koh Y, et al. Prognostic value of metabolic tumour volume on baseline ¹⁸F-FDG PET/CT in addition to NCCN-IPI in patients with diffuse large B-cell lymphoma: further stratification of the group with a high-risk NCCN-IPI. *Eur J Nucl Med Mol Imaging*. 2019;46(7):1417-1427.
42. Hans CP, Weisenburger DD, Greiner TC, et al. Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. *Blood*. 2004;103(1):275-282.
43. 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(7):1142-1154.
44. Cheson BD, Pfistner B, Juweid ME, et al; International Harmonization Project on Lymphoma. Revised response criteria for malignant lymphoma. *J Clin Oncol*. 2007;25(5):579-586.

45. Harrell FE Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15(4):361-387.
46. 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(5):793-808.
47. Hothorn T, Hornik K, Zeileis A. Unbiased recursive partitioning: a conditional inference framework. *J Comput Graph Stat*. 2006;15(3):651-674.
48. Senjo H, Hirata K, Izumiyama K, et al. High metabolic heterogeneity on baseline 18F-FDG PET/CT predicts worse prognosis of newly diagnosed diffuse large B-cell lymphoma. *Blood*. 2019;134(Suppl 1):488.
49. 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(6):1334-1344.
50. Cottreau AS, Hapdey S, Chartier L, et al. Baseline total metabolic tumor volume measured with fixed or different adaptive thresholding methods equally predicts outcome in peripheral T cell lymphoma. *J Nucl Med*. 2017;58(2):276-281.
51. Meignan M, Cottreau AS, Versari A, 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(30):3618-3626.
52. Kim TM, Paeng JC, Chun IK, et al. Total lesion glycolysis in positron emission tomography is a better predictor of outcome than the International Prognostic Index for patients with diffuse large B cell lymphoma. *Cancer*. 2013;119(6):1195-1202.
53. Adams HJA, de Klerk JMH, Fijnheer R, et al. Prognostic superiority of the National Comprehensive Cancer Network International Prognostic Index over pretreatment whole-body volumetric-metabolic FDG-PET/CT metrics in diffuse large B-cell lymphoma. *Eur J Haematol*. 2015;94(6):532-539.
54. Nanni C, Cottreau AS, Lopci E, et al. Report of the 6th International Workshop on PET in lymphoma. *Leuk Lymphoma*. 2017;58(10):2298-2303.
55. Cottreau AS, Buvat I, Kanoun S, et al. Is there an optimal method for measuring baseline metabolic tumor volume in diffuse large B cell lymphoma? *Eur J Nucl Med Mol Imaging*. 2018;45(8):1463-1464.
56. Barrington SF, Meignan M. Time to prepare for risk adaptation in lymphoma by standardizing measurement of metabolic tumor burden. *J Nucl Med*. 2019; 60(8):1096-1102.
57. Meignan M. Quantitative FDG-PET: a new biomarker in PMBCL. *Blood*. 2015;126(8):924-926.