

Combination of baseline metabolic tumour volume and early response on PET/CT improves progression-free survival prediction in DLBCL

N. George Mikhaeel¹ · Daniel Smith¹ · Joel T. Dunn² · Michael Phillips² · Henrik Møller³ · Paul A. Fields⁴ · David Wrench⁴ · Sally F. Barrington²

Received: 6 August 2015 / Accepted: 27 November 2015 / Published online: 23 February 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract

Background The study objectives were to assess the prognostic value of quantitative PET and to test whether combining baseline metabolic tumour burden with early PET response could improve predictive power in DLBCL.

Methods A total of 147 patients with DLBCL underwent FDG-PET/CT scans before and after two cycles of RCHOP. Quantitative parameters including metabolic tumour volume (MTV) and

total lesion glycolysis (TLG) were measured, as well as the percentage change in these parameters. Cox regression analysis was used to test the relationship between progression-free survival (PFS) and the study variables. Receiver operator characteristics (ROC) analysis determined the optimal cut-off for quantitative variables, and Kaplan–Meier survival analysis was performed.

Results The median follow-up was 3.8 years. As MTV and TLG measures correlated strongly, only MTV measures were used for multivariate analysis (MVA). Baseline MTV (MTV-0) was the only statistically significant predictor of PFS on MVA. The optimal cut-off for MTV-0 was 396 cm³. A model combining MTV-0 and Deauville score (DS) separated the population into three distinct prognostic groups: good (MTV-0 < 400; 5-year PFS > 90 %), intermediate (MTV-0 ≥ 400+ DS1-3; 5-year PFS 58.5 %) and poor (MTV-0 ≥ 400+ DS4-5; 5-year PFS 29.7 %).
Conclusions MTV-0 is an important prognostic factor in DLBCL. Combining MTV-0 and early PET/CT response improves the predictive power of interim PET and defines a poor-prognosis group in whom most of the events occur.

✉ N. George Mikhaeel
george.mikhaeel@gstt.nhs.uk

Daniel Smith
daniel.smith@gstt.nhs.uk

Joel T. Dunn
joel.dunn@kcl.ac.uk

Michael Phillips
michael.phillips@kcl.ac.uk

Henrik Møller
Henrik.moller@kcl.ac.uk

Paul A. Fields
paul.fields@gstt.nhs.uk

David Wrench
david.wrench@gstt.nhs.uk

Sally F. Barrington
sally.barrington@kcl.ac.uk

¹ Department of Clinical Oncology, Guy's and St Thomas' NHS Foundation Trust, Westminster Bridge Road, SE1 7EH London, UK

² PET Imaging Centre at St Thomas' Hospital, Division of Imaging Sciences and Biomedical Engineering, King's College London, London, UK

³ Department of Cancer Epidemiology and Population Health, King's College London, London, UK

⁴ Department of Haematology, Guy's and St Thomas' NHS Foundation Trust, London, UK

Keywords Positron emission tomography · Diffuse large B-cell lymphoma · Metabolic tumour volume · Deauville score · Response assessment

Introduction

The cure rate of diffuse large B-cell lymphoma (DLBCL) has improved over the last two decades with the addition of rituximab to CHOP chemotherapy (RCHOP) and improvements in dose intensity and supportive care [1, 2]. Nonetheless, a proportion of patients are not cured with RCHOP, either due to primary refractory disease or later relapse following an initial response. Salvage treatments after RCHOP are less effective than after CHOP, probably because the increased cure rate

with upfront treatment leaves a worse prognostic group for salvage [3, 4]. Early identification of patients unlikely to be cured with RCHOP is an important step towards testing alternative approaches in order to improve their chance of cure.

¹⁸F-fluorodeoxyglucose (FDG) PET/CT performed early during a course of chemotherapy (interim PET, iPET) shows early response and has been found to be prognostic in Hodgkin lymphoma [5–9]. This has led to several clinical studies testing response-adapted treatment algorithms. However in DLBCL, iPET studies have shown conflicting results [10–21]. Even in studies showing a significant difference between early complete metabolic response (CMR) and partial metabolic response (PMR), the progression-free survival (PFS) in the partial response group has remained too high, around 50 % at 2–5 years, to consider early change of treatment. More recent response criteria, e.g. Deauville score (DS) and the reduction in the maximum SUV (Δ SUV_{max}), have not increased the positive predictive value of iPET in DLBCL, and there is a need to improve the current response criteria by exploring additional parameters to improve the predictive ability of interim PET.

In addition to early treatment response, baseline characteristics of the lymphoma, including tumour burden, impact significantly on outcome [22]. Quantitative imaging methods using PET/CT to assess total tumour burden are under development and may be more predictive than existing methods [23]. We hypothesised that patients may achieve a good early metabolic response but their chance of cure may be different based on pre-treatment tumour burden. The objectives of this study were to assess the prognostic value of quantitative PET measurements, particularly metabolic tumour burden, and to test whether combining measurements of metabolic tumour burden at baseline with early PET response could improve the prognostic ability of iPET in DLBCL.

Patients and methods

Study population

This is a retrospective study including 147 consecutive patients treated at Guy's and St Thomas' Hospital, London, between March 2005 and August 2012, according to a prospectively agreed-upon departmental protocol. The study was approved by the hospital review committee, and all patients gave written informed consent. The inclusion criteria were as follows: histologically proven DLBCL, treatment with RCHOP chemotherapy, PET/CT scans before and after two cycles of chemotherapy, and a minimum follow-up of 12 months. Exclusion criteria were coexistent low-grade lymphoma, other malignancy or active sarcoidosis, previous anthracycline chemotherapy, and no assessable disease on baseline PET/CT.

Treatment protocol

All patients received standard-dose RCHOP chemotherapy [1]. Figure 1 summarises the treatment algorithm. PET/CT scan after two cycles (iPET2) was not used to change treatment unless there was definite evidence of progression or lack of response, taking into consideration all clinical and laboratory information. Patients with CMR continued to a total of 6 cycles of chemotherapy or 3–4 cycles followed by involved-field radiotherapy in the case of stage I or II non-bulky disease. Patients with PMR after two cycles received two further cycles of RCHOP followed by repeat PET/CT scan (iPET4). If iPET4 showed CMR or further response with residual uptake (PMR), treatment continued to six cycles. All patients had an end-of-treatment scan (CT if there was CMR on iPET; PET/CT if there was PMR on iPET). Consolidation radiotherapy was administered according to departmental protocol and clinician discretion.

In total, seven patients changed treatment due to progressive disease while on RCHOP. Two patients progressed after cycle 2: one patient with rising lactate dehydrogenase (LDH) and no significant response on PET, and a second patient showing progression of disease on PET. Both were changed to salvage chemotherapy but progressed and died. Two other patients progressed after three cycles: one with recurrence of palpable bulky lymphadenopathy and another with cytologically proven CNS relapse. Both were changed to salvage chemotherapy but progressed and died. Two patients progressed after cycle 4, both with increased FDG uptake: one was successfully salvaged with further chemotherapy and radiotherapy, and one died of progressive disease following salvage high-dose chemotherapy and stem cell transplant. Finally, one patient progressed after cycle 5 with increasing disease bulk on CT, which was biopsy-proven, and died following salvage chemotherapy.

With regard to radiotherapy (RT), 50 patients had RT. In 35 patients, RT was planned treatment as part of a combined treatment approach for stage I or II disease (relapse=1/35). Of the 15 with stage III or IV disease who received RT, three had stage III disease and received RT to a residual focus of uptake (relapse=0/3), and 12 had stage IV disease and received RT for the following reasons: site of bony disease (two), contralateral testicular RT (one), residual activity on PET-6 (four), initial bulk and DS 3 on PET-6 (two), initial bulk and residual mass (one), clinician discretion (two) (relapse=1/12).

PET/CT scanning

PET/CT scans were performed after a 6-h fast. Images were acquired from skull base to upper thighs 90 minutes after administration of 370 MBq of ¹⁸F-FDG using DST or VCT scanners (GE Healthcare, Waukesha, WI, USA). Baseline and

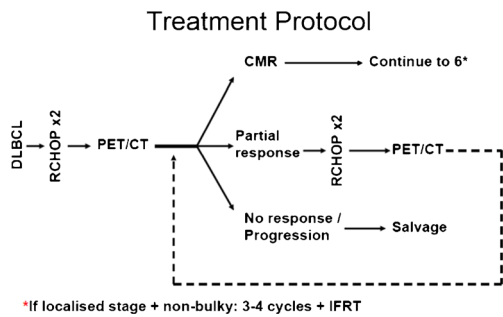


Fig. 1 Treatment protocol. CMR = complete metabolic response (defined as Deauville score 1–3). Partial response was defined as Deauville score 4–5 with improvement compared to baseline. No response/progression was defined as Deauville score 4–5 with no improvement or progression compared to baseline

interim scans were performed on the same scanner. Quality control procedures to ensure accurate quantitation developed for clinical trials have been previously described [24]. Paired scans were displayed using a fixed SUV scale and colour table, and were reviewed by an experienced nuclear medicine physician (SFB), who was blinded to clinical outcome and other baseline parameters such as the International Prognostic Index (IPI). The Deauville score was allotted by comparing the uptake in the residual lesion with the highest uptake, if present, with the intensity of uptake in normal mediastinum and liver. A score of 5 was given if the maximum SUV in the tumour was ≥ 3 times greater than the maximum SUV in a large region of normal liver.

In-house software developed by MP was used to automatically segment tumour volumes with $SUV \geq 2.5$. Volumes were edited by SFB to remove physiological uptake, e.g. myocardium adjacent to mediastinal nodal disease. All tumour volumes were summed to calculate the total metabolic tumour volume (MTV). Tumour lesion glycolysis (TLG) was calculated as $MTV \times \text{mean SUV in the volume}$.

The following parameters were determined on the baseline scan: $SUV_{\text{max-0}}$, MTV_{0} and TLG_{0} . The following parameters were determined on the scan after two cycles: $SUV_{\text{max-2}}$, MTV_{2} , TLG_{2} and the Deauville score. Changes in these parameters during treatment were also calculated: % change in SUV_{max} (ΔSUV_{max}), % change in MTV (ΔMTV), % change in TLG (ΔTLG).

Statistical analysis

The endpoint of the study was PFS, defined as the time from diagnosis to the point of progression or death from any cause. Patients still alive were censored at the date of last contact.

To determine which measures were predictive of PFS, univariate Cox regression (CR) analysis was performed on all quantitative parameters as well as IPI and Deauville score. The IPI score was analysed according to the standard prognostic groups: low risk (0–1), low intermediate (2), high

intermediate (3), and high (4–5). Continuous measures were grouped into tertiles. Variables which were significantly associated with PFS in univariate analysis were included in multivariate analysis in order to identify measures independently predictive of PFS.

Receiver operator characteristics (ROC) analysis was performed on continuous variables to determine optimal cut-off values. The effect of individual parameters on PFS was studied with Kaplan–Meier (KM) survival analysis using dichotomous grouping based on ROC analysis. IPI scores were grouped into 0–1 and 2–5, and DS scores into 1–3 (CMR) and 4–5 (incomplete response). For ΔSUV_{max} , we used the previously reported cut-off of 66 % [22, 28].

Based on these results, a prognostic model combining baseline and response parameters was constructed on which KM survival analysis was performed.

Statistical significance was considered at $p < 0.05$. Data were analysed using SPSS software (IBM Corp. Released 2011. IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY: IBM Corp.).

Results

The median follow-up for 147 patients was 3.8 years (range 1.3–7.9 years). The clinical characteristics are summarised in Table 1. The 5-year PFS and OS rates for the whole group were 65.4 % and 73.7 %, respectively. Table 2 shows the descriptive statistics for the quantitative PET parameters.

Univariate CR analysis (Table 3) showed that the following variables were statistically significant: IPI, MTV_{0} , TLG_{0} , DS, MTV_{2} , TLG_{2} , $SUV_{\text{max-2}}$ and ΔSUV_{max} .

MTV and TLG measures correlated very strongly at both baseline and after two cycles (Pearson's correlation coefficient, $r = 0.916$ and $r = 0.961$, respectively), as tertile groups were almost identical. Therefore, only MTV was included in multivariate analysis (MVA). The MVA included IPI, DS, MTV_{0} , MTV_{2} , $SUV_{\text{max-2}}$ and ΔSUV_{max} , and showed that MTV_{0} was the only independent statistically significant measure.

ROC analysis (Fig. 2) showed that the optimal cut-off for sensitivity and specificity for MTV_{0} was 396 cm^3 (sensitivity 0.88, specificity 0.62) and for TLG_{0} was 4541 (sensitivity 0.76, specificity 0.63). The optimal cut-offs for MTV_{2} and TLG_{2} were skewed towards very low values (1.8 and 5.6 respectively), reflecting the fact that most patients had significant reduction in FDG activity after two cycles of chemotherapy. Similarly, the cut-offs for ΔMTV and ΔTLG were 99.76 % and 99.9 %. Therefore, only the cut-offs for MTV_{0} and TLG_{0} were considered clinically meaningful.

We also examined the relationship between the presence of bulky disease (defined as maximum tumour dimension

Table 1 Patient characteristics

Patient characteristics		<i>n</i> = 147 (%)
Sex	Female	74
	Male	73
Age	Range	22–86
	Median	57
	≥60 years	71 (48 %)
Performance status	0–1	103 (70 %)
	2–4	44 (30 %)
Raised LDH		93 (63 %)
Extra-nodal sites ≥2		73 (50 %)
Bulky disease (≥10 cm)		59 (40 %)
Stage	I	17 (11 %)
	II	29 (20 %)
	III	16 (11 %)
	IV	85 (58 %)
IPI	0/1	45 (31 %)
	2	18 (12 %)
	3	38 (26 %)
	4/5	46 (31 %)

IPI = International Prognostic Index, LDH = lactate dehydrogenase

(MTD) in any direction ≥10 cm as defined by ROC analysis, data not shown) and MTV-0. Overall, 59 patients (40 %) had bulky disease. As expected, the presence of bulky disease correlated with high MTV-0 (parametric Pearson's correlation coefficient, $r=0.557$, $p<0.001$). However, MTV-0 predicted the outcome of patients better than MTD. Overall concordance was 79 % (41 % had low bulk and low MTV-0, and 38 % had high bulk and high MTV-0). Thirty-one patients (21 %) had discordance of bulk and MTV-0 classification. Twenty-six patients (18 %) had low bulk but high MTV-0, and their 5-year PFS was 45 %. Five patients (3 %) had high bulk but low MTV-0, and their 5-year PFS was 100 % (Fig. 3).

Table 2 Mean, median and range of quantitative PET parameters

	Mean	Median	Min	Max
MTV-0 (cm ³)	990.14	595.12	1.50	7357.20
TLG-0 (cm ³)	6815.91	4669.52	5.69	36,570.00
SUVmax-0	27.89	27.25	5.38	110.52
MTV-2 (cm ³)	34.97	0.59	0.00	1608.70
TLG-2 (cm ³)	213.45	1.64	0.00	13,135.00
SUVmax-2	2.05	2.73	0.00	10.34
ΔMTV (%)	−95.81	−99.95	−100.00	−6.95
ΔTLG (%)	−96.88	−99.98	−100.00	−31.71
ΔSUVmax (%)	−69.03	−70.21	−100.00	41.39

KM survival analysis showed that the following variables were associated with PFS: IPI (0–1 v 2–5: $X^2=14.17$, $p<0.001$), MTV-0 (<400 vs. >400: $X^2=34.17$, $p<0.001$), TLG-0 (<4500 vs. >4500: $X^2=20.13$, $p<0.001$), DS (1–3 vs. 4–5: $X^2=10.51$, $p=0.001$), MTV-2 (<1.8 vs. >1.8: $X^2=11.32$, $p=0.001$), TLG-2 (<5.6 vs. >5.6: $X^2=10.36$, $p=0.001$), SUVmax-2 (<3 vs. >3: $X^2=5.3$, $p=0.021$), ΔMTV (>99.8 % vs. <99.8 %: $X^2=6.36$, $p=0.012$), ΔTLG (>99.9 % vs. <99.9 %: $X^2=7.43$, $p=0.006$), ΔSUVmax (>66 % vs. <66 %: $X^2=13.68$, $p<0.001$). Figure 4 shows examples of KM PFS analysis.

For overall survival analysis, the corresponding values were: IPI (0–1 vs. 2–5: $X^2=18.4$, $p<0.001$), MTV-0 (<400 vs. >400: $X^2=23.9$, $p<0.001$), TLG-0 (<4500 vs. >4500: $X^2=12.8$, $p<0.001$), DS (1–3 vs. 4–5: $X^2=7.13$, $p=0.008$), MTV-2 (<1.8 vs. >1.8: $X^2=4.1$, $p=0.043$), TLG-2 (<5.6 vs. >5.6: $X^2=3.65$, $p=0.056$), SUVmax-2 (<3 vs. >3: $X^2=2.79$, $p=0.095$), ΔMTV (>99.8 % vs. <99.8 %: $X^2=3.06$, $p=0.08$), ΔTLG (>99.9 % vs. <99.9 %: $X^2=2.81$, $p=0.094$), ΔSUVmax (>66 % vs. <66 %: $X^2=6.49$, $p=0.011$).

To test the study hypothesis that a combination of baseline metabolic tumour burden and response after two cycles can improve the prognostic ability of iPET-2 in DLBCL, we constructed a prognostic model combining MTV-0—the only statistically significant baseline measure on MVA—and DS. The aim was to identify a group with sufficiently poor prognosis that clinicians might consider changing treatment, and large enough to include most of the population events. Patients were classified into four groups according to the two parameters: MTV-0<400+DS 1–3 (46 patients, 31 %), MTV-0<400+DS 4–5 (20 patients, 14 %), MTV-0≥400+DS 1–3 (36 patients, 24 %) and MTV-0≥400+DS 4–5 (45 patients, 31 %) (Fig. 5a).

This model showed that patients with high baseline tumour burden (MTV-0≥400) and poor response (DS 4–5) had a much worse prognosis than the remainder of the study population with 5-year PFS 29.7 %. Patients with low MTV-0 had a favourable prognosis regardless of the DS after two cycles (5-year PFS 90.9 and 95.0 %), and patients with high MTV-0 who responded well according to DS had an intermediate prognosis (5-year PFS 58.5 %). The worst prognosis group included 31 % of the study population but 58 % of the total events (29/50). Hazard ratios (compared to the group with low MTV-0 and low DS) were 0.49 (low MTV-0+high DS, 95 % CI 0.06–4.21), 4.7 (high MTV-0+low DS, 95 % CI 1.7–13.0) and 10.1 (high MTV-0+high DS, 95 % CI 3.9–26.2).

We examined other models—e.g. replacing MTV-0 with TLG-0 or IPI, replacing DS with ΔSUVmax or adding other parameters to MTV-0 and DS—but none was able to predict a worse prognostic group, and the current model remained the simplest and most predictive. The model combining MTV-0 and ΔSUVmax, for example, is shown in Fig. 5b. The worst prognosis group included only 11 % of patients and 26 % of the events.

Table 3 Cox regression analysis

	LEVELS	CASES (n = 147)	UNIVARIATE			MULTIVARIATE		
			HR	95 % CI for HR		HR	95 % CI for HR	
IPI groups	(0,1)	45	1.00			1.00		
	2	18	4.23	1.34	13.34	2.95	0.82	10.60
	3	38	4.74	1.73	12.98	2.30	0.74	7.21
	(4,5)	46	5.49	2.07	14.53	2.98	0.98	9.08
	<i>TREND:</i>		$\chi^2 = 12.39$	<i>P-value = .0004</i>		$\chi^2 = 2.78$	<i>P-value = .0955</i>	
DS	1	34	1.00			1.00		
	2	18	0.74	.234	2.321	0.21	0.04	1.09
	3	30	0.49	.169	1.402	0.06	0.01	0.42
	4	47	1.33	.621	2.844	0.09	0.01	0.62
	5	18	3.86	1.725	8.640	0.23	0.03	1.88
<i>TREND:</i>		$\chi^2 = 7.61$	<i>P-value = .0058</i>		$\chi^2 = 0.95$	<i>P-value = .3303</i>		
SUVmax-0 Tertiles	Lower	49	1.00					
	Middle	49	1.44	.74	2.79			
	Upper	49	0.86	.41	1.78			
<i>TREND:</i>		$\chi^2 = 0.17$	<i>P-value = .6831</i>					
MTV-0 Tertiles	Lower	49	1.00			1.00		
	Middle	49	3.77	1.49	9.51	2.73	.89	8.40
	Upper	49	5.81	2.38	14.14	3.46	1.10	10.86
<i>TREND:</i>		$\chi^2 = 16.54$	<i>P-value = .0000</i>		$\chi^2 = 4.00$	<i>P-value = .0454</i>		
TLG-0 Tertiles	Lower	49	1.00					
	Middle	49	2.96	1.24	7.10			
	Upper	49	4.90	2.11	11.38			
<i>TREND:</i>		$\chi^2 = 14.95$	<i>P-value = .0001</i>					
SUVmax-2 Tertiles	Lower	49	1.00			1.00		
	Middle	49	1.40	.65	2.99	2.56	.20	33.22
	Upper	49	2.46	1.22	4.96	0.98	.06	16.10
<i>TREND:</i>		$\chi^2 = 6.72$	<i>P-value = .0096</i>		$\chi^2 = 2.40$	<i>P-value = .1216</i>		
MTV-2 Tertiles	Lower	59	1.00			1.00		
	Middle	39	1.49	.68	3.27	4.16	.49	35.29
	Upper	49	3.09	1.58	6.05	8.08	.72	90.67
<i>TREND:</i>		$\chi^2 = 11.21$	<i>P-value = .0008</i>		$\chi^2 = 3.02$	<i>P-value = .0820</i>		
TLG-2 Tertiles	Lower	59	1.00					
	Middle	39	1.49	.68	3.27			
	Upper	49	3.09	1.58	6.05			
<i>TREND:</i>		$\chi^2 = 11.21$	<i>P-value = .0008</i>					
Δ SUVmax Tertiles	Lower	49	1.00			1.00		
	Middle	49	0.97	.46	2.07	1.04	.32	3.31
	Upper	49	2.16	1.11	4.21	1.16	.32	4.17
<i>TREND:</i>		$\chi^2 = 5.42$	<i>P-value = .0199</i>		$\chi^2 = 0.09$	<i>P-value = .7701</i>		
Δ MTV Tertiles	Lower	59	1.00					
	Middle	39	2.63	1.30	5.33			
	Upper	49	2.02	.99	4.12			
<i>TREND:</i>		$\chi^2 = 3.83$	<i>P-value = .0502</i>					
Δ TLG Tertiles	Lower	59	1.00					
	Middle	39	2.66	1.31	5.40			
	Upper	49	2.00	.98	4.07			
<i>TREND:</i>		$\chi^2 = 3.71$	<i>P-value = .0542</i>					

Discussion

The addition of rituximab to CHOP chemotherapy has improved the outcome of first-line treatment for DLBCL [1, 2], but salvage therapy for patients who are not cured with RCHOP is less effective than that after CHOP [3, 4]. Early identification of patients who are unlikely to be cured with RCHOP is an important step to enable testing of alternative

approaches. Although early response assessment using iPET has been established in Hodgkin lymphoma, its role remains controversial in DLBCL due to conflicting results [10–21] and the inability to detect a group with sufficiently poor prognosis even with new response criteria such as DS or Δ SUVmax [25]. For example, a recently published study using DS showed a statistically significant difference between scores 1–3 and 4–5 (85 % vs. 72 % respectively, p=0.047), but the

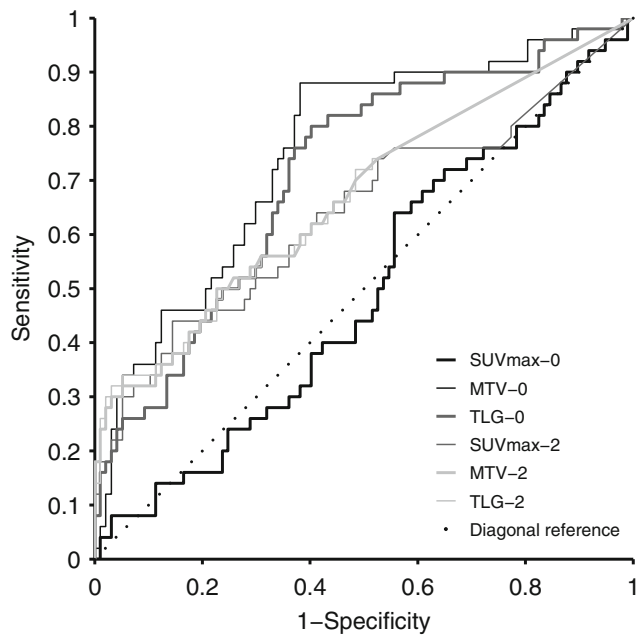


Fig. 2 Receiver operator characteristics analysis

worse group still had excellent PFS [18]. Another study using Δ SUVmax also showed a statistically significant difference between the two groups, while the worse prognostic group still had a high PFS [17].

In order to improve the predictive ability of PET, we assessed the prognostic value of various quantitative PET parameters, and found that the pre-treatment total metabolic burden was the strongest predictor of PFS. MTV-0 and TLG-0 were the most predictive quantitative measures on univariate analysis and were superior to the

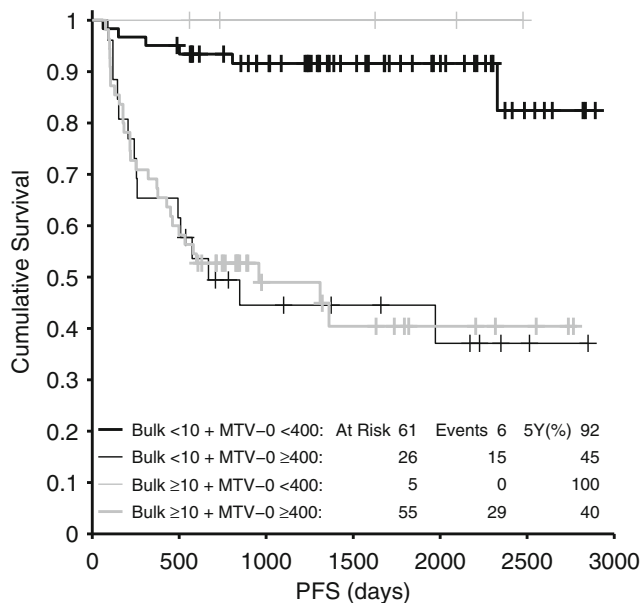


Fig. 3 Relationship between initial bulk and MTV-0

standard tools for assessing pre-treatment risk and early response, namely the IPI and DS. As results using MTV and TLG were almost identical, we included only MTV parameters in MVA, which showed that MTV-0 remained statistically significant while IPI and DS did not. These results confirm previous studies showing the prognostic value of baseline metabolic tumour burden [26–29]; however, this is the first study to show that pre-treatment total metabolic burden is more prognostic than either early response or IPI in DLBCL.

We also examined the relationship between the presence of bulky disease (as defined by anatomical size criteria) and MTV-0. Bulk has traditionally been considered a poor prognostic factor, although the IPI did not include bulky disease [22]. The prognostic value of bulk was demonstrated in the rituximab era, and the optimal cut-off appeared to be 10 cm (measuring the maximum tumour dimension=MTD) in RCHOP-treated patients [30]. Measuring MTD is technically easier and quicker than measuring MTV, so we compared the two parameters to determine which was more predictive and whether bulk expressed as MTD could predict the same outcome while being easier to measure. Using MTD cut-off of ≥ 10 cm, 59 patients (40 %) were found to have bulky disease. As expected, the presence of bulky disease correlated with high MTV-0; however, MTV-0 predicted the outcome of patients much better than MTD. Overall, the two parameters agreed in 79 % of patients, but in 21 % of patients who had discordance of bulk and MTV-0 classification, MTV-0 correctly predicted the outcome regardless of the presence or absence of bulk. Twenty-six patients (18 %) had low bulk but high MTV-0, and their 5-year PFS was 45 %, while five patients (3 %) had high bulk but low MTV-0, and their 5-year PFS was 100 %.

While baseline metabolic burden was a very strong predictor in this population, the change in metabolic burden was less predictive. The percentage reductions in MTV and TLG after two cycles (Δ MTV and Δ TLG, respectively) were borderline statistically significant on univariate analysis. On ROC analysis, the optimal cut-offs for Δ MTV and Δ TLG were skewed (99.76 % and 99.9 %, respectively), indicating that most of the population had a significant reduction in metabolic volume after two cycles of chemotherapy. Similarly, the optimal cut-offs for MTV-2 and TLG-2 were skewed towards very low values (1.8 and 5.6, respectively).

On the other hand, ROC analysis showed that the optimal cut-off for sensitivity and specificity for MTV-0 was 396 cm³ (rounded to 400) and for TLG-2 was 4541 cm³ (rounded to 4500). These were considered clinically meaningful and were included in further analysis. Using these cut-offs, KM survival analysis

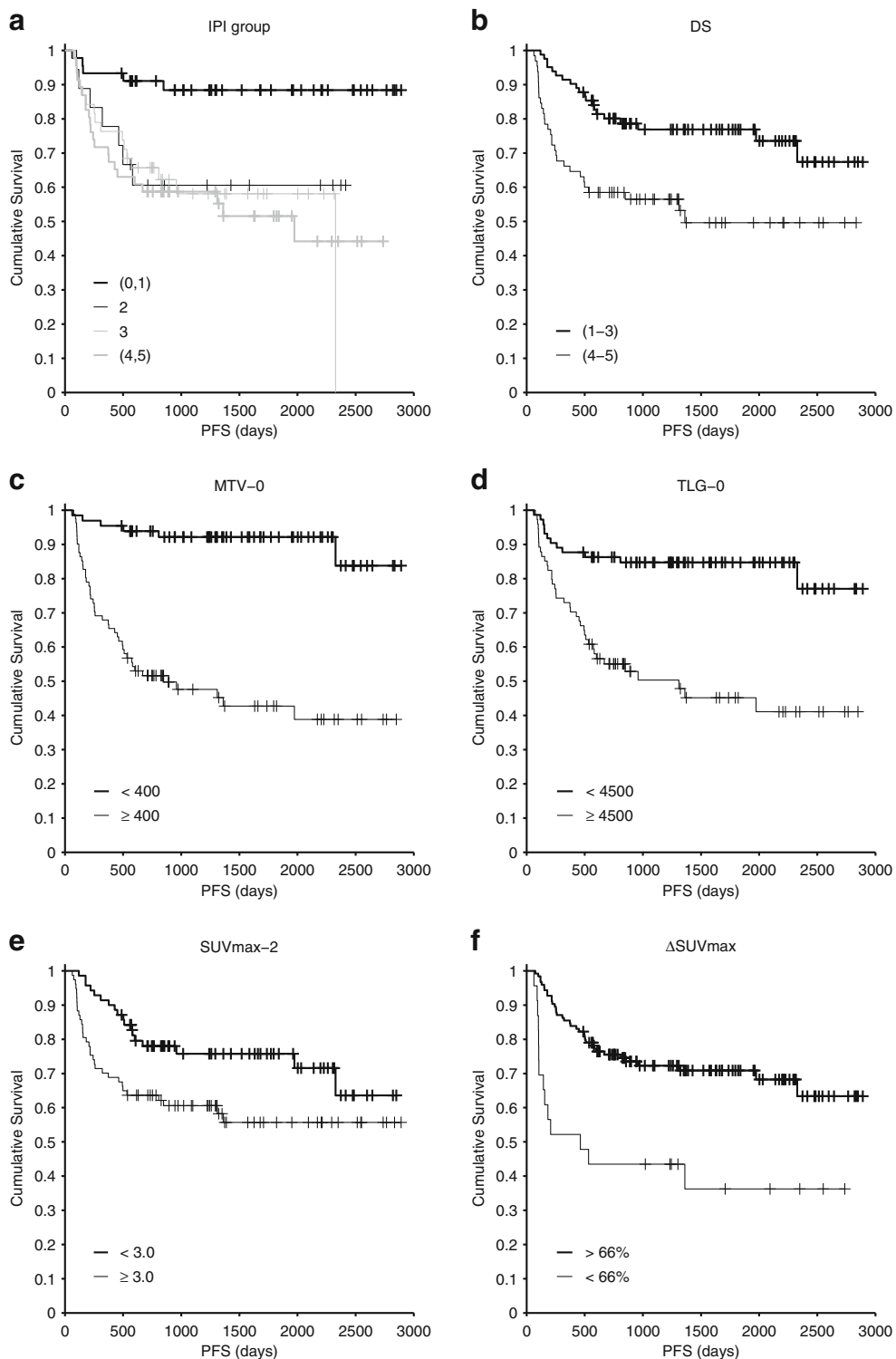


Fig. 4 Kaplan–Meier progression-free survival curves. **a:** IPI groups **b:** DS **c:** MTV-0 **d:** TLG-0 **e:** SUVmax-2 **f:** ΔSUVmax

separated the study population into two distinct prognostic groups with statistically significant differences in PFS.

Other studies have reported optimal cut-offs for MTV-0 ranging from 220 to 625 cm³ [26, 27, 29]. The optimal

cut-off value is dependent on the characteristics of the study population and the methodology of MTV measurement. Studies with populations with better prognosis (e.g. earlier stage, less bulky disease and/or low IPI) resulted in lower cut-offs and vice versa. Song et al. [26] studied

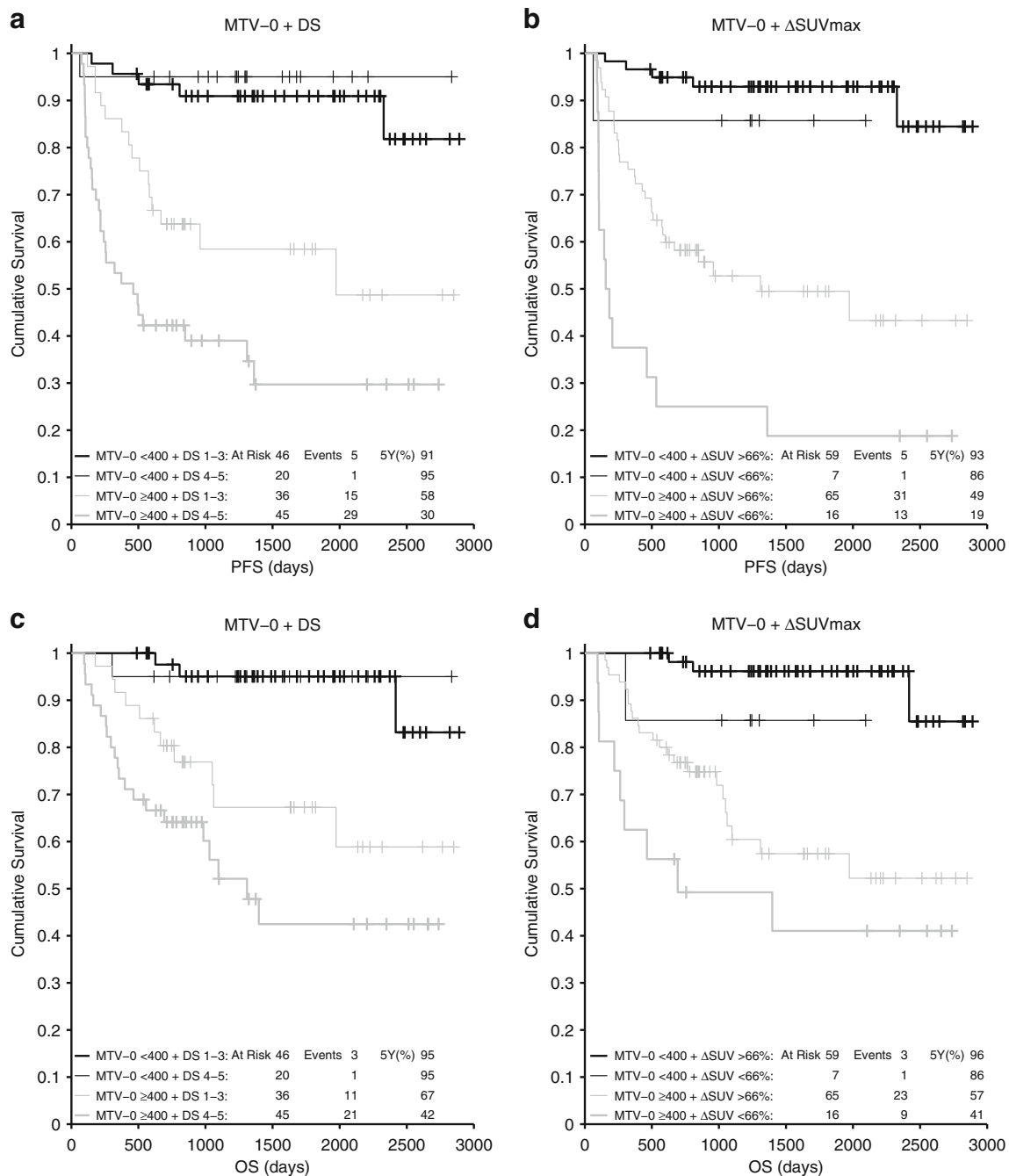


Fig. 5 Prognostic model **a**: Model combining MTV-0 + DS (PFS). **b**: Model combining MTV-0 + ΔSUVmax (PFS). **c**: Model combining MTV-0 + DS (OS). **d**: Model combining MTV-0 + ΔSUVmax (OS)

only stage 2–3 patients without extra-nodal involvement and few (4.1 %) with bulky disease (>5 cm), and reported the optimal cut-off as 220 cm³. Sasanelli et al. [29] included patients from eight institutions, most with stage 3–4 (82 %), with bulky disease (>10 cm) in 36 % and two or more extra-nodal sites in 32 %, and found an optimal cut-off of 550 cm³. Our study was non-selective and included consecutive patients from one institution with all

stages and presentations of disease, and so is likely more representative of the general population of DLBCL in clinical practice. Interestingly, the optimal cut-off in our study was in the middle of the wide reported range. The other factor affecting MTV-0 cut-off is the measurement method. We applied the method used by Song et al. [26], defining the MTV by SUVmax cut-off of 2.5, while Sasanelli [29] and Casanovas [27] used a threshold of 41 % of SUVmax

in each lesion to define MTV. Meignan et al. [31] recently published a study evaluating different methods using phantom and patient data, and concluded that 41 % is the optimal cut-off to define MTV in lymphomas, in line with solid tumours [32]. Our study, however, was conducted before this recent publication. Validation studies comparing the prognostic ability of the two methods are needed to resolve this issue, but it seems that in the current study population, the MTV-0 cut-off of 400 cm³, defined by SUVmax > 2.5, resulted in a clinically significant separation of low- and high-MTV-0 groups (3-year PFS 92.2 % vs. 47.6 %) and compares favourably to MTV-0 defined by 41 % of SUVmax (3-year PFS 77 % vs. 60 %) [29].

One of the study objectives was to test the hypothesis that a combination of baseline measurement of metabolic tumour burden and early response assessment could improve the prognostic ability of iPET-2 in DLBCL. The model combining MTV-0 and DS was able to separate significantly different prognostic groups and to identify a group with a prognosis sufficiently poor that clinicians might consider changing treatment. Patients with high baseline tumour burden (MTV-0 ≥ 400) and poor response after two cycles (DS 4–5) had a much worse prognosis than the remainder of the study population with 5-year PFS of 29.7 %. This compares favourably with either DS alone (5-year PFS of poor prognosis group 49.6 %) or ΔSUVmax alone (5-year PFS of poor prognosis group 36.2 %, but this group included only 16 % of patients). Patients with low MTV-0 had a favourable prognosis regardless of their DS after two cycles (5-year PFS 90.9 and 95.0 %), indicating that pre-treatment prognosis is more important than early response in this group. Conversely, patients with high MTV-0 who responded well according to DS had an intermediate prognosis (5-year PFS 58.5 %). It is also worth noting that MTV-0 separated the DS 4–5 patients into the best and the worst prognostic groups, and correctly reclassified 20 patients as good prognosis with 5-year PFS 95 %. Finally, the worst prognosis group included 58 % of the total number of events (29/50) and 31 % of patients, making this model useful for identifying patients with poor prognosis.

Our results have significant clinical implications, and the proposed prognostic model based on both initial metabolic burden and response after two cycles of chemotherapy could be used in both clinical practice and future clinical trials. Most notably, the separation of poor early responders (DS 4–5) into very different prognostic groups (best and worst) depending on MTV-0 can potentially be used to test response-adapted treatments. However, the proposed model must be validated in another independent data set to confirm its prognostic

value and optimal cut-off, and to test it against MTV defined by 41 % cut-off.

The recently published imaging guidelines (ICML guidelines) [23] and Lugano classification [33] now recommend performing PET/CT for staging in all FDG-avid lymphomas, and baseline scan information should be readily available. In addition, efforts to standardise PET methods to ensure quantitative accuracy in everyday practice [32, 34] and the increasing availability of automated software from manufacturers to measure metabolic volumes make quantitative analysis in the clinic a realistic option in the near future.

Conclusions

Baseline metabolic tumour volume is an important prognostic factor in diffuse large B-cell lymphoma treated with RCHOP chemotherapy. The combination of baseline metabolic tumour volume and response on PET/CT after two cycles measured by Deauville score improves the predictive power of interim PET and separates patients into distinct groups with different prognosis, defining a poor-prognosis group which includes most of the events after treatment.

IPI, international prognostic index; *DS*, Deauville score after 2 cycles; *SUVmax-0*, baseline maximum standardised uptake value; *MTV-0*, baseline metabolic tumour volume; *TLG-0*, baseline total lesion glycolysis; *SUVmax-2*, maximum standardised uptake value after 2 cycles; *MTV-2*, metabolic tumour volume after 2 cycles; *TLG-2*, total lesion glycolysis after 2 cycles; *ΔSUVmax*, percentage change in maximum standardised uptake value; *ΔMTV*, percentage change in metabolic tumour volume; *ΔTLG*, percentage change in total lesion glycolysis; *ROC*, receiver operator characteristics

Compliance with ethical standards

Author contributions NGM, SFB and DS designed the study. MP designed the in-house software. SFB reviewed the scans. DS collected the clinical data. JTD and HM performed statistical analysis. NGM, PAF and DW recruited patients. All contributed to final analysis and writing of the manuscript.

Funding There was no external funding for this study.

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

Ethical approval This article does not contain any studies with animals performed by any of the authors.

All procedures performed were in accordance with the ethical standards of the institutional research committee and with the 1964 Declaration of Helsinki and its later amendments.

Informed consent was obtained from all individual participants included in the study.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Coiffier B, Lepage E, Briere J, et al. CHOP chemotherapy plus rituximab compared with CHOP alone in elderly patients with diffuse large-B-cell lymphoma. *N Engl J Med*. 2002;346:235–42.
2. Pfreundschuh M, Kuhnt E, Trumper L, et al. CHOP-like chemotherapy with or without rituximab in young patients with good-prognosis diffuse large-B-cell lymphoma: 6-year results of an open-label randomised study of the MabThera International Trial (MInT) Group. *Lancet Oncol*. 2011;12:1013–22.
3. Gisselbrecht C, Glass B, Mounier N, et al. Salvage regimens with autologous transplantation for relapsed large B-cell lymphoma in rituximab era. *J Clin Oncol*. 2010;28(27):4184–90.
4. Friedberg JW. Relapsed/Refractory Diffuse Large B-cell Lymphoma. *Hematology*. 2011;2011:498–505.
5. Hutchings M, Mikhaeel NG, Fields PA, et al. Prognostic value of interim FDG-PET after two or three cycles of chemotherapy in Hodgkin lymphoma. *Ann Oncol*. 2005;16:1160–8.
6. Gallamini A, Hutchings M, Rigacci L, et al. Early interim 2- F-18 Fluoro-2-Deoxy-D-Glucose positron emission tomography is prognostically superior to international prognostic score in advanced-stage Hodgkin's lymphoma: A report from a joint Italian-Danish study. *J Clin Oncol*. 2007;25:3746–52.
7. Cerci JJ, Pracchia LF, Linardi CCG, et al. (18)F-FDG PET After 2 Cycles of ABVD Predicts Event-Free Survival in Early and Advanced Hodgkin Lymphoma. *J Nucl Med*. 2010;51:1337–43.
8. Zinzani PL, Rigacci L, Stefoni V, et al. Early interim 18F-FDG PET in Hodgkin's lymphoma: evaluation on 304 patients. *Eur J Nucl Med Mol Imaging*. 2012;39:4–12.
9. Gallamini A, Barrington SF, Biggi A, et al. The predictive role of interim positron emission tomography for Hodgkin lymphoma treatment outcome is confirmed using the interpretation criteria of the Deauville five-point scale. *Haematologica*. 2014;99(6):1107–13.
10. Mikhaeel NG, Timothy AR, O'Doherty MJ, et al. 18-FDG-PET as a prognostic indicator in the treatment of aggressive Non-Hodgkin's Lymphoma-comparison with CT. *Leuk Lymphoma*. 2000;39:543–53.
11. Mikhaeel NG, Hutchings M, Fields PA, et al. FDG-PET after two to three cycles of chemotherapy predicts progression-free and overall survival in high-grade non-Hodgkin lymphoma. *Ann Oncol*. 2005;16:1514–23.
12. Haioun C, Itti E, Rahmouni A, et al. F-18 fluoro-2-deoxy-D-glucose positron emission tomography (FDG-PET) in aggressive lymphoma: an early prognostic tool for predicting patient outcome. *Blood*. 2005;106:1376–81.
13. 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;50:527–33.
14. Yang D-H, Min J-J, Song H-C, et al. Prognostic significance of interim (18)F-FDG PET/CT after three or four cycles of R-CHOP chemotherapy in the treatment of diffuse large B-cell lymphoma. *Eur J Cancer*. 2011;47:1312–8.
15. Zinzani PL, Gandolfi L, Broccoli A, et al. Midtreatment 18F-fluorodeoxyglucose positron-emission tomography in aggressive non-Hodgkin lymphoma. *Cancer*. 2011;117:1010–8.
16. Cashen AF, Dehdashti F, Luo J, et al. (18)F-FDG PET/CT for Early Response Assessment in Diffuse Large B-Cell Lymphoma: Poor Predictive Value of International Harmonization Project Interpretation. *J Nucl Med*. 2011;52:386–92.
17. Casasnovas RO, Meignan M, Berriolo-Riedinger A, et al. SUVmax reduction improves early prognosis value of interim positron emission tomography scans in diffuse large B-cell lymphoma. *Blood*. 2011;118:37–43.
18. Pregno P, Chiappella A, Bello 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;119:2066–73.
19. 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;30:184–90.
20. 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 DeltaSUVmax. *Eur J Nucl Med Mol Imaging*. 2013;40:1312–20.
21. Barrington SF, Mikhaeel NG. When should FDG-PET be used in the modern management of lymphoma? *Br J Haematol*. 2013;164(3):315–28. doi:10.1111/bjh.12601.
22. A predictive model for aggressive non-Hodgkin's lymphoma. The International Non-Hodgkin's Lymphoma Prognostic Factors Project. *N Engl J Med*. 1993; 329:987–94.
23. 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):3048–58.
24. Barrington SF, MacKewn JE, Schleyer P, et al. Establishment of a UK-wide network to facilitate the acquisition of quality assured FDG-PET data for clinical trials in lymphoma. *Ann Oncol*. 2011;22:739–45.
25. Moskowitz CH. Interim PET-CT in the management of diffuse large B-cell lymphoma. *Hematology*. 2012;2012(1):397–401.
26. Song MK, Chung JS, Shin HJ, 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.
27. Casasnovas RO, Sasanelli M, Berriolo-Riedinger A, et al. Baseline metabolic tumour volume is predictive of patient outcome in diffuse large B cell lymphoma. *Blood*. 2012;120(21):abstract 1598.
28. 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;119:1195–202.
29. 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):2017–22. doi:10.1007/s00259-014-2822-7.
30. Pfreundschuh M, Ho AD, Cavallin-Stahl E, et al. Prognostic significance of maximum tumour (bulk) diameter in young patients with good-prognosis diffuse large-B-cell lymphoma treated with CHOP-like chemotherapy with or without rituximab: an exploratory analysis of the MabThera International Trial Group (MInT) study. *Lancet Oncol*. 2008;9:435–44.
31. Meignan M, Sasanelli M, Casasnovas RO, 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.
32. Boellaard R, O'Doherty MJ, Weber WA, et al. FDG PET and PET/CT: EANM procedure guidelines for tumour PET imaging: version 1.0. *Eur J Nucl Med Mol Imaging*. 2010;37:181–200.
 33. Cheson BD, Fisher RI, Barrington SF, et al. Recommendations for initial evaluation, staging, and response assessment of Hodgkin and non-Hodgkin lymphoma: the Lugano classification. *J Clin Oncol*. 2014;32(27):3059–68.
 34. <http://earl.eanm.org/cms/website.php>