Metabolic Tumor Volume for Response Prediction in Advanced-Stage Hodgkin Lymphoma

Jasmin Mettler[1], Horst Müller[2], Conrad-Amadeus Voltin[1], Christian Baues[3], Bernd Klaeser[4,5], Alden Moccia[5,6], Peter Borchmann[7], Andreas Engert[7], Georg Kuhnert[1], Alexander Drzezga[1], Markus Dietlein[1], and Carsten Kobe[1]

[1]Department of Nuclear Medicine, University Hospital of Cologne, Cologne, Germany; [2]German Hodgkin Study Group, Department of Internal Medicine I, University Hospital of Cologne, Cologne, Germany; [3]Department of Radiation Oncology, University Hospital of Cologne, Cologne, Germany; [4]Department of Nuclear Medicine, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland; [5]Swiss Group for Clinical Cancer Research, Bern, Switzerland; [6]Department of Medical Oncology, Oncology Institute of Southern Switzerland, Bellinzona, Switzerland; and [7]Department of Internal Medicine I, University Hospital of Cologne, Cologne, Germany

^{18}F-FDG PET/CT for staging Hodgkin lymphoma may allow for accurate and reliable assessment of the metabolic tumor volume (MTV) as a baseline risk factor. Our aim was to analyze the prognostic impact of MTV measurements obtained by different means in advanced-stage Hodgkin lymphoma patients treated within the German Hodgkin Study Group HD18 trial. **Methods:** Within HD18, 310 patients underwent ^{18}F-FDG PET/CT scanning for staging, which was available to the central review panel for quantitative analysis. We calculated the MTV by 4 different thresholding methods and performed receiver-operating-characteristic analysis to evaluate the potential for prediction of early response determined by PET after 2 cycles (PET-2) of dose-escalated bleomycin, etoposide, doxorubicin, cyclophosphamide, vincristine, procarbazine, and prednisone (eBEACOPP). Logistic regression was used to evaluate its prognostic value concerning progression-free survival and overall survival. **Results:** All of the different MTV calculations predicted PET-2 response to a moderate and comparable degree (area under the curve, 0.62–0.63; P = 0.01–0.06). With none of the measuring methods did the receiver-operating-characteristic curves point to any unique cutoffs; rather, a wide range of possible cutoffs was indicated. None of the MTV measurements was prognostic for progression-free survival (hazard ratio, 1.2–1.5; P = 0.15–0.52) or overall survival (hazard ratio, 1.0–1.5; P = 0.95–0.27). **Conclusion:** Baseline MTV as determined by different means is a predictive factor for early response to eBEACOPP after 2 cycles. However, value as a prognostic factor after a highly effective PET-2–adapted treatment strategy could not be observed.

Key Words: PET/CT; Hodgkin lymphoma; metabolic tumor volume; response prediction; progression-free survival; overall survival

O ver the last 20–30 y, therapy of advanced-stage Hodgkin lymphoma has evolved to the point at which treatment offers a more than 90% likelihood of success (1). ^{18}F-FDG PET/CT has become an established tool for staging and response assessment in Hodgkin lymphoma patients (2,3). The PET-directed personalized adaption of chemo- and radiotherapy has become reality and improved the outcome for many patients with this malignancy (4–8).

However, individualized, risk-stratified treatment might be further improved because PET performed for initial staging may contain more information than currently used in clinical routine. In Hodgkin lymphoma, it was suggested that the initial tumor volume might have a prognostic impact (9). Nowadays, convenient methods have been introduced for the measurement of metabolic tumor volume (MTV) using PET, and first analyses have brought promising results for patients treated with doxorubicin, bleomycin, vinblastine, and dacarbazine (ABVD) (10–14).

Here, we raised the question of how and to what extent initial risk stratification within the group of advanced-stage Hodgkin lymphoma patients treated with dose-escalated bleomycin, etoposide, doxorubicin, cyclophosphamide, vincristine, procarbazine, and prednisone (eBEACOPP) can be further improved by calculation of MTV.

Our aim was therefore to analyze the predictive and prognostic role of MTV at baseline in terms of PET response after 2 cycles of chemotherapy, progression-free survival (PFS), and overall survival (OS). As there are various approaches to the assessment of MTV, applying absolute, relative, or adaptive thresholding methods, we used 4 common methods to determine MTV (15).

MATERIALS AND METHODS

Patients

Between May 2008 and July 2014, 2,101 newly diagnosed patients, aged 18–60 y, were recruited for the German Hodgkin Study Group HD18 trial. HD18 is a phase III multicenter randomized study for advanced stages of Hodgkin lymphoma, including Ann Arbor stages IIB (with a large mediastinal mass or extranodal involvement as risk factors), IIIA/B, and IVA/B.

^{18}F-FDG PET scans at baseline and their central review were optional in this trial. ^{18}F-FDG PET scans were performed and available for central review for 429 patients at baseline. Since determination of MTV relies on absolute quantification of activity, the analysis set was restricted to those scans that were available in digital form, that contained CT data for anatomic correlation and attenuation correction, and that included information for SUV calculation (e.g., body weight). Thus, 310 scans remained suitable for quantitative analyses (Fig. 1).
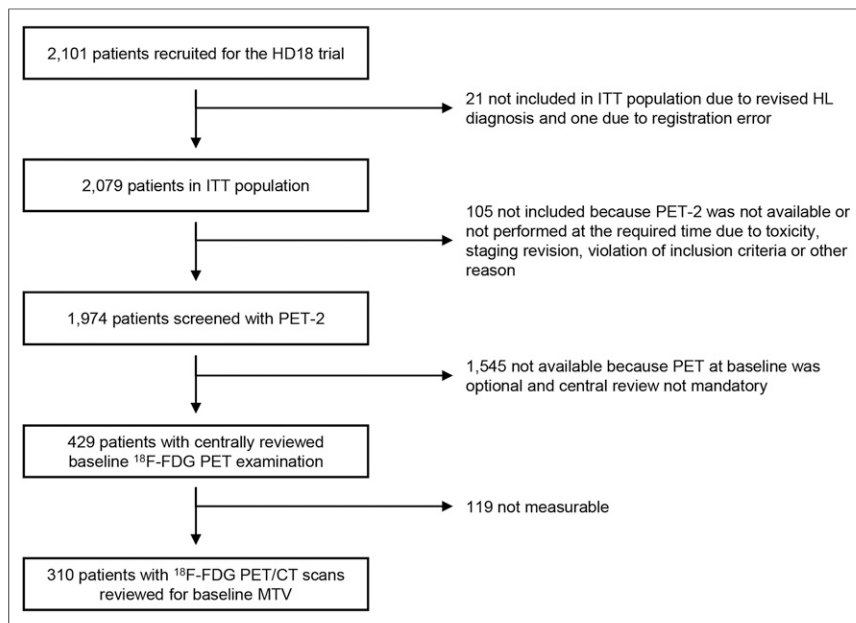
FIGURE 1. Of 2,101 patients recruited, baseline ^{18}F-FDG PET was centrally reviewed in 429, of whom 310 were suitable for quantitative analyses. ITT = intention to treat; HL = Hodgkin lymphoma.

The study was performed in accordance with the Declaration of Helsinki and the International Conference on Harmonization guidelines for good clinical practice and registered at ClinicalTrials.gov (NCT00515554). The institutional review board approved this study, and all subjects gave written informed consent.

Image Analyses

We analyzed MTV in all available baseline PET scans using the Beth Israel plugin for FIJI (ImageJ), a shareware from the Beth Israel Deaconess Medical Center, Division of Nuclear Medicine and Molecular Imaging. In a previous study, the Beth Israel plugin was successfully tested for comparability in terms of MTV determination with commercial software (*16*).

Every site of lymphoma tissue was manually delineated by an experienced nuclear medicine physician. To avoid interobserver variations in our multicenter trial, all MTV calculations were performed by the same reviewer. If lymphoma sites had heterogeneous uptake, regions of very high uptake were delineated separately as an individual volume

TABLE 1
Patient Characteristics

Characteristic	Patients with centrally reviewed ^{18}F-FDG PET at baseline			Patients without ^{18}F-FDG PET at baseline
	Measured (*n* = 310)	Not measurable (*n* = 119)	Total (*n* = 429)	(*n* = 1,545)
Mean age at HL diagnosis (y)	34 ± 11 (range, 18–60)	34 ± 12 (range, 18–59)	34 ± 11 (range, 18–60)	34 ± 12 (range, 18–60)
Ann Arbor stage IV	126 (41%)	42 (35%)	168 (39%)	547 (35%)
Large mediastinal mass*	92 (30%)	41 (35%)	133 (31%)	439 (28%)
Extranodal involvement	64 (21%)	32 (27%)	96 (22%)	309 (20%)
High ESR†	269 (87%)	107 (90%)	376 (88%)	1,325 (86%)
≥3 lymph nodes	165 (54%)	69 (58%)	234 (55%)	1,020 (66%)
B symptoms	176/308 (57%)	70/118 (59%)	246/426 (58%)	1,006/1,542 (65%)
IPS				
0–1	107/306 (35%)	39/118 (33%)	146/424 (34%)	460/1,527 (30%)
2–3	157/306 (51%)	64/118 (54%)	221/424 (52%)	817/1,527 (54%)
4–7	42/306 (14%)	15/118 (13%)	57/424 (13%)	250/1,527 (16%)
PET-2, ^{18}F-FDG uptake‡				
DS1	103 (33%)	48 (40%)	151 (35%)	532 (34%)
DS2	46 (15%)	21 (18%)	67 (16%)	262 (17%)
DS3	79 (26%)	23 (19%)	102 (24%)	379 (25%)
DS4	81 (26%)	27 (23%)	108 (25%)	372 (24%)

*at least one third of maximal thoracic diameter as measured on chest radiography.
†≥50 mm/h without B symptoms or ≥30 mm/h with B symptoms.
‡Analysis set contained no patient with Deauville score of 5 (*17*).
HL = Hodgkin lymphoma; ESR = erythrocyte sedimentation rate; IPS = International Prognostic Score; DS = Deauville score.
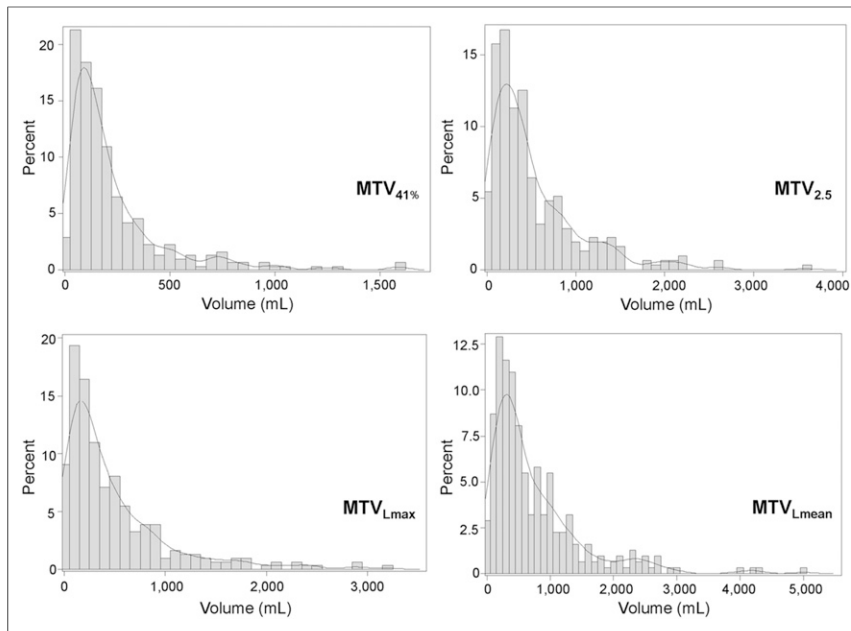Data are *n* followed by percentage in parentheses, except for age.

FIGURE 2. Histograms of MTV distribution assessed by 4 different thresholding methods: $MTV_{41\%}$, $MTV_{2.5}$, MTV_{Lmax}, and MTV_{Lmean}.

of interest to prevent underestimation of MTV when using relative thresholds.

Extranodal lymphoma sites such as lung, liver, and bone marrow were individually delineated if they showed focal hypermetabolism. Homogeneous patterns of increased uptake within bone marrow were not included. Spleen involvement was given if focal hypermetabolism was found or in cases of diffuse hypermetabolism if the uptake could be clearly seen to exceed the liver background.

First, the SUV_{mean} and the SUV_{max} were obtained in the background-region liver by manually drawing a spheric volume of interest 3 cm in diameter, preferably within the right lobe.

For MTV measurements, we drew volumes of interest around every site of lymphoma tissue with increased ^{18}F-FDG uptake. MTV was obtained as the sum of supra-threshold voxels of all lymphoma lesions. All MTV calculations were performed using the following thresholding methods: 41% of the SUV_{max} within the respective lymphoma site ($MTV_{41\%}$), SUV_{max} of liver background (MTV_{Lmax}), SUV_{mean} of liver background (MTV_{Lmean}), and a fixed SUV of 2.5 ($MTV_{2.5}$).

All HD18-recruited patients underwent 2 cycles of eBEACOPP followed by an early interim PET examination (PET-2). PET-2 was interpreted visually by the central review panel using the Deauville score (*17,18*).

TABLE 2
Pearson Correlations of MTV Measures

Parameter	$MTV_{41\%}$	$MTV_{2.5}$	MTV_{Lmax}	MTV_{Lmean}
$MTV_{41\%}$	1.00	0.78	0.69	0.65
$MTV_{2.5}$	0.78	1.00	0.87	0.86
MTV_{Lmax}	0.69	0.87	1.00	0.97
MTV_{Lmean}	0.65	0.86	0.97	1.00

Statistics

Patient characteristics were recorded using descriptive statistics, and MTV distributions were visualized in histograms. Thresholding methods were compared by Pearson product moment correlation coefficients. Receiver-operating-characteristic analysis was performed to evaluate baseline MTV as a predictor of PET-2 response, using the visual liver uptake as cutoff (Deauville score, 4) (*17,18*). Additionally, the differences in log-transformed MTVs between the PET-2–positive and –negative groups were tested for significance with *t* tests (pooled or Satterthwaite). Prognostic impact was analyzed using Cox proportional hazards regression of PFS and OS on the log-transformed MTVs and compared with the International Prognostic Score. PFS was defined as the time from completion of staging until progression, relapse, or death from any cause or until the day when information was last received on the patient's disease status. OS was defined as the time from completion of staging until death from any cause or until the last day that the patient was known to be alive. All statistical computations were performed using SAS, version 9.4 (SAS Institute).

RESULTS

Patients

Patient characteristics are shown in Table 1. The mean age of the 310 patients was 34 y (range, 18–60 y). At baseline, 92 patients (30%) had a large mediastinal mass. Extranodal involvement occurred in 64 patients (21%). In 269 patients (87%), a high erythrocyte sedimentation rate was a risk factor. In 165 patients (54%), more than 3 lymph node areas were involved. In 176 of 308 patients (57%), there were B symptoms. The International Prognostic Score was 0–1 in 107 of 306 patients (35%), 2–3 in 157 (51%), and 4–7 in 42 (14%).

MTV

Histograms of the MTV distributions are presented in Fig. 2 and show a right-skewed distribution for all thresholding methods.

The median values of $MTV_{41\%}$, $MTV_{2.5}$, MTV_{Lmax}, and MTV_{Lmean} were 142 mL (range, 6–1,590 mL), 355 mL (range, 4–3,563 mL), 297 mL (range, 2–3,192 mL), and 483 mL (range, 7–4,995 mL).

Correlations among absolute thresholding methods ranged from 0.86 to 0.97, and correlations between relative and absolute thresholding methods ranged from 0.65 to 0.78 (Table 2).

Predictive Power for PET-2 Positivity

The receiver-operating-characteristic curves for PET-2 response, derived from MTV using different thresholding methods, show comparable curve progressions (SUV_{max}) FIG. 3. Areas under the curve for $MTV_{41\%}$, $MTV_{2.5}$, MTV_{Lmax}, and MTV_{Lmean} were 0.63 ($P = 0.06$), 0.62 ($P = 0.04$), 0.63 ($P = 0.01$), and 0.62 ($P = 0.02$), respectively. The differences between the log-transformed MTVs of PET-2–positive and –negative groups were all significant (pooled or Satterthwaite *t* tests with $P < 0.0001$ to $P = 0.0011$).

Progression-Free and Overall Survival

Univariate hazard ratios for prognostication of PFS with $MTV_{41\%}$, $MTV_{2.5}$, MTV_{Lmax}, and MTV_{Lmean} were 1.2 (95% confidence interval, 0.70–2.04; $P = 0.52$), 1.5 (95% confidence interval, 0.89–2.39;
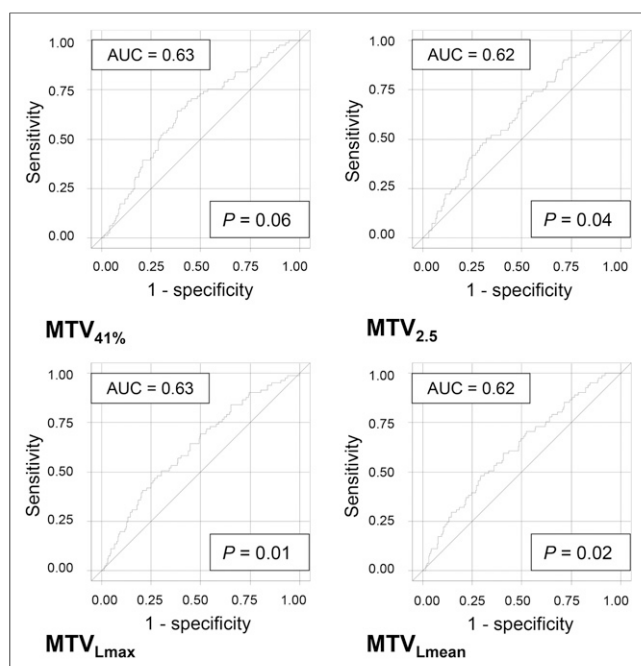
FIGURE 3. Receiver-operating-characteristic curves of MTV distribution assessed by 4 different thresholding methods: $MTV_{41\%}$, $MTV_{2.5}$, MTV_{Lmax}, and MTV_{Lmean}. (81/310 = 26% PET-positive patients with Deauville score ≥ 4.)

$P = 0.14$), 1.4 (95% confidence interval, 0.86–2.11; $P = 0.19$), and 1.5 (95% confidence interval, 0.88–2.43; $P = 0.15$), respectively. None of the measuring methods for MTV was prognostic for PFS or OS (Table 3). The multivariate analysis on the right side of Table 3 shows that the International Prognostic Score, conversely, was significantly prognostic for both PFS and OS ($P < 0.01$). In multivariate analysis testing the MTVs adjusting for the International Prognostic Score the hazard ratios of the MTVs approximated 1.0, the point of indifference ($P = 0.40$–0.79).

DISCUSSION

The following findings emerge from our analysis of MTV in 310 patients with advanced-stage Hodgkin lymphoma.

First, each method used for calculation of the MTV works comparably and equally well in terms of prediction of PET response after 2 cycles of eBEACOPP. With none of the measuring methods did the receiver-operating-characteristic curves point to any unique cutoffs. Instead, they indicated a wide range of possible cutoffs.

Second, baseline MTV failed to predict PFS and OS in the subset of advanced-stage Hodgkin lymphoma patients undergoing PET response–adapted treatment with eBEACOPP followed by PET-guided radiotherapy. Hence, individual PET-2 response assessment cannot be omitted for further therapy adaption.

MTV as Prognostic Factor

Various studies have indicated the prognostic potential of baseline MTV in Hodgkin lymphoma patients (*10–14*). Cottereau et al. showed that MTV is prognostic of PFS and OS in 258 patients with early-stage Hodgkin lymphoma who received a standard combined-modality treatment (*11*). Song et al. showed that a certain MTV cutoff at baseline could identify a subgroup of early-stage Hodgkin lymphoma patients treated with ABVD at a higher risk for progression, who might benefit from additional therapy such as irradiation (*13*). Kanoun et al. showed baseline MTV to be superior to tumor bulk (≥10 cm) in prognosticating progression-free and disease-specific survival in 59 Hodgkin lymphoma patients treated with anthracycline-based chemotherapy (*10*). In 267 Hodgkin lymphoma patients with stage I and II disease with and without risk factors, Akhtari et al. observed that MTV prognosticates both PFS and OS (*14*). The authors concluded that MTV might improve outcome if incorporated into initial risk stratification.

The studies presented by Cottereau, Song, Kanoun, and Akhtari had in common that patients had been treated with ABVD. Conversely, our analysis set was restricted to advanced-stage patients who had been

TABLE 3
Univariate and Multivariate Cox Regressions of PFS (16 Events) and OS (7 Events) on Log-Transformed
MTVs Compared with IPS

	MTV						
	Univariate analysis			Multivariate analysis with IPS			
Parameter	Hazard ratio	95% CI	*P*	Hazard ratio	95% CI	*P*	IPS (*P*)
PFS							
$MTV_{41\%}$	1.2	0.70–2.04	0.52	0.9	0.53–1.62	0.79	0.0022
$MTV_{2.5}$	1.5	0.89–2.39	0.14	1.2	0.75–1.86	0.48	0.0053
MTV_{Lmax}	1.4	0.86–2.11	0.19	1.1	0.70–1.73	0.69	0.0048
MTV_{Lmean}	1.5	0.88–2.43	0.15	1.1	0.66–1.89	0.69	0.0061
OS							
$MTV_{41\%}$	1.0	0.46–2.27	0.95	0.7	0.31–1.60	0.40	0.0006
$MTV_{2.5}$	1.5	0.72–3.25	0.27	1.1	0.60–2.05	0.48	0.0018
MTV_{Lmax}	1.2	0.64–2.34	0.54	0.9	0.46–1.68	0.69	0.0010
MTV_{Lmean}	1.2	0.58–2.47	0.62	0.8	0.37–1.61	0.48	0.0008

IPS = International Prognostic Score; CI = confidence interval.

treated with eBEACOPP, a chemotherapy regimen proven to provide a higher complete remission rate and longer PFS than ABVD (*1*).

However, our analysis is limited by the low number of events under PET-adapted eBEACOPP for advanced-stage patients. Hence, results can be applied only to populations undergoing chemotherapy similar to the regimen we used. Furthermore, the high efficiency of PET-adapted eBEACOPP might have concealed the prognostic value of MTV that could be observed for patients receiving ABVD treatment.

Methods of MTV Measurement

To date, MTV measurement is not a standardized procedure (*19*). Approaches range from use of fixed absolute thresholds to use of adaptive methods based, for example, on signal-to-background ratio (*15,16,19*). Here, MTV was analyzed using absolute thresholds and a relative threshold of 41% of SUV_{max}. In 106 patients with peripheral T-cell lymphoma, Cottereau et al. found baseline MTV to be a relevant risk factor regardless of the thresholding method used (*15*). Different methods of MTV measurement have already been compared in Hodgkin lymphoma. Significantly higher MTVs were found at staging using a fixed SUV cutoff of 2.5 than using a relative threshold of 41% of SUV_{max} (*16*). Accordingly, we found that the absolute value for total MTV can diverge heavily depending on the method used for calculation (Fig. 2). The highest absolute values were calculated for MTV_{Lmean}, followed by $MTV_{2.5}$, MTV_{Lmax}, and $MTV_{41\%}$. However, we confirmed that every method used in our study led to a similar result concerning the prediction of response at PET-2. For future trials, we would therefore recommend selecting parameters that are readily available and promise high reproducibility between different institutions. For some authors, $MTV_{41\%}$ appears to be the most reproducible option, and it is also recommended by the European Association of Nuclear Medicine (*19,20*). When considering the use of MTV as a foundation for treatment decisions, it is necessary to ensure that standardization is possible, especially with the next generation of scanners about to be launched (*21*).

CONCLUSION

MTV can predict response after 2 cycles of eBEACOPP independently of the method of determination used. Because PFS and OS could not be appraised in advanced-stage patients undergoing PET-2–adapted treatment with eBEACOPP, omitting PET after 2 cycles and at the end of chemotherapy to deescalate treatment is not an option in this setting.

DISCLOSURE

ACKNOWLEDGMENTS

REFERENCES

1. Skoetz N, Will A, Monsef I, Brillant C, Engert A, von Tresckow B. Comparison of first-line chemotherapy including escalated BEACOPP versus chemotherapy including ABVD for people with early unfavourable or advanced stage Hodgkin lymphoma. *Cochrane Database Syst Rev.* 2017;5:CD007941.
2. 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;32:3059–3068.
3. El-Galaly TC, d'Amore F, Mylam KJ, et al. Routine bone marrow biopsy has little or no therapeutic consequence for positron emission tomography/computed tomography-staged treatment-naive patients with Hodgkin lymphoma. *J Clin Oncol.* 2012;30:4508–4514.
4. Engert A, Haverkamp H, Kobe C, et al. Reduced-intensity chemotherapy and PET-guided radiotherapy in patients with advanced stage Hodgkin's lymphoma (HD15 trial): a randomised, open-label, phase 3 non-inferiority trial. *Lancet.* 2012;379:1791–1799.
5. Borchmann P, Haverkamp H, Lohri A, et al. Progression-free survival of early interim PET-positive patients with advanced stage Hodgkin's lymphoma treated with BEACOPP escalated alone or in combination with rituximab (HD18): an open-label, international, randomised phase 3 study by the German Hodgkin Study Group. *Lancet Oncol.* 2017;18:454–463.
6. Johnson P, Federico M, Kirkwood A, et al. Adapted treatment guided by interim PET-CT scan in advanced Hodgkin's lymphoma. *N Engl J Med.* 2016;374:2419–2429.
7. André MP, Girinsky T, Federico M, et al. Early positron emission tomography response-adapted treatment in stage I and II Hodgkin lymphoma: final results of the randomized EORTC/LYSA/FIL H10 trial. *J Clin Oncol.* 2017;35:1786–1794.
8. Barrington SF, Johnson PWM. ¹⁸F-FDG PET/CT in lymphoma: has imaging directed personalized medicine become a reality? *J Nucl Med.* 2017;58:1539–1544.
9. Gobbi PG, Valentino F, Bassi E, et al. Chemoresistance as a function of the pretherapy tumor burden and the chemotherapy regimen administered: differences observed with 2 current chemotherapy regimens for advanced Hodgkin lymphoma. *Clin Lymphoma Myeloma Leuk.* 2011;11:396–402.
10. 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;41:1735–1743.
11. Cottereau AS, Versari A, Loft A, et al. Prognostic value of baseline metabolic tumor volume in early-stage Hodgkin lymphoma in the standard arm of the H10 trial. *Blood.* 2018;131:1456–1463.
12. Moskowitz AJ, Schöder H, Gavane S, et al. Prognostic significance of baseline metabolic tumor volume in relapsed and refractory Hodgkin lymphoma. *Blood.* 2017;130:2196–2203.
13. Song MK, Chung JS, Lee JJ, et al. Metabolic tumor volume by positron emission tomography/computed tomography as a clinical parameter to determine therapeutic modality for early stage Hodgkin's lymphoma. *Cancer Sci.* 2013;104:1656–1661.
14. Akhtari M, Milgrom SA, Pinnix CC, et al. Re-classifying patients with early-stage Hodgkin lymphoma based on functional radiographic markers at presentation. *Blood.* 2018;131:84–94.
15. Cottereau 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;58:276–281.
16. Kanoun S, Tal I, Berriolo-Riedinger A, et al. Influence of software tool and methodological aspects of total metabolic tumor volume calculation on baseline [¹⁸F] FDG PET to predict survival in Hodgkin lymphoma. *PLoS One.* 2015;10:e0140830.
17. Meignan M, Gallamini A, Haioun C. Report on the first international workshop on interim-PET-scan in lymphoma. *Leuk Lymphoma.* 2009;50:1257–1260.
18. 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;99:1107–1113.
19. 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;41:1113–1122.
20. 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.
21. Kuhnert G, Boellaard R, Sterzer S, et al. Impact of PET/CT image reconstruction methods and liver uptake normalization strategies on quantitative image analysis. *Eur J Nucl Med Mol Imaging.* 2016;43:249–258.